



A repetitive elements perspective in Polycomb epigenetics

Valentina Casa^{1,2} and Davide Gabellini^{1*}

¹ Division of Regenerative Medicine, Stem Cells, and Gene Therapy, Dulbecco Telethon Institute and San Raffaele Scientific Institute, Milano, Italy

² Università Vita-Salute San Raffaele, Milano, Italy

Edited by:

Michèle Amouyal, CNRS, France

Reviewed by:

Igor Kovalchuk, University of

Lethbridge, Canada

Michèle Amouyal, CNRS, France

*Correspondence:

Davide Gabellini, Division of
Regenerative Medicine, Stem Cells,
and Gene Therapy, Dulbecco
Telethon Institute and San Raffaele
Scientific Institute, DIBIT 2, 5A3-44,
Via Olgettina 58, 20132 Milano, Italy.
e-mail: gabellini.davide@hsr.it

Repetitive elements comprise over two-thirds of the human genome. For a long time, these elements have received little attention since they were considered non-functional. On the contrary, recent evidence indicates that they play central roles in genome integrity, gene expression, and disease. Indeed, repeats display meiotic instability associated with disease and are located within common fragile sites, which are hotspots of chromosome re-arrangements in tumors. Moreover, a variety of diseases have been associated with aberrant transcription of repetitive elements. Overall this indicates that appropriate regulation of repetitive elements' activity is fundamental. Polycomb group (PcG) proteins are epigenetic regulators that are essential for the normal development of multicellular organisms. Mammalian PcG proteins are involved in fundamental processes, such as cellular memory, cell proliferation, genomic imprinting, X-inactivation, and cancer development. PcG proteins can convey their activity through long-distance interactions also on different chromosomes. This indicates that the 3D organization of PcG proteins contributes significantly to their function. However, it is still unclear how these complex mechanisms are orchestrated and which role PcG proteins play in the multi-level organization of gene regulation. Intriguingly, the greatest proportion of Polycomb-mediated chromatin modifications is located in genomic repeats and it has been suggested that they could provide a binding platform for Polycomb proteins. Here, these lines of evidence are woven together to discuss how repetitive elements could contribute to chromatin organization in the 3D nuclear space.

Keywords: repeats, Polycomb, non-protein-coding RNA, nuclear structure, FSHD muscular dystrophy

INTRODUCTION

In the last decade, when the genomic sequences of *Homo sapiens* and several model organisms became available, there was the realization that the number of protein-coding genes does not correlate with organism complexity. In fact, worms or flies have approximately the same number of protein-coding genes as mice or humans (Taft et al., 2007). On the other hand, the non-protein coding component of the genomic DNA, and in particular repetitive elements, represent a progressively larger proportion of the genome in organisms with increasing complexity (Neguembor and Gabellini, 2010). Recent estimations indicate that repetitive sequences could account for up to 66–69% of the human genome (De Koning et al., 2011). While this strongly suggests that it might significantly contribute to higher eukaryotes sophistication, the repetitive fraction of the genome is largely ignored.

The advent of next generation sequencing (NGS) has permitted a genome-wide view to gene expression and chromatin structure. However, NGS-based studies often take into account only reads for which a unique genomic alignment can be obtained, thus discarding data deriving from repetitive DNA (Myers et al., 2011). Despite this, there is increasing evidence of the peculiar functions of the repeated (epi)genome. For example, the role of DNA repeats in chromosome structural organization, gene regulation, genome integrity, and evolution has been described (Kidwell and Lisch, 2000; Lander et al., 2001; Waterston et al., 2002; Feschotte, 2008; Ting et al., 2011; Zhu et al., 2011).

DNA repeats can be also transcribed, frequently in a cell and tissue-specific fashion. Analyses based on Cap Analysis of Gene Expression (CAGE) technology from the Functional Annotation of Mouse (FANTOM) project, revealed an unexpectedly large proportion of capped-transcripts initiating from repetitive units. It has been suggested that these can provide regulatory elements to protein-coding genes, such as alternative promoters, exons, or polyadenylation sites, and ncRNAs, thus significantly expanding the regulatory capability of higher eukaryote genomes (Wang et al., 2007; Bourque et al., 2008; Faulkner et al., 2009; Tyekucheva et al., 2011). Moreover, binding sites for important regulatory factors such as CTCF or TP53 are often associated with genomic repeats (Wang et al., 2007; Bourque et al., 2008; Chadwick, 2008; Simeonova et al., 2012).

Repetitive elements can either mobilize or rearrange in somatic tissues, thus providing an unexpected dynamic dimension to the normal physiology of the soma, but also contributing to the etiopathogenesis of diseases (Kazazian et al., 1988; Ting et al., 2011; Zhu et al., 2011). For the role they can play in genome plasticity, repeats need to be finely tuned. To accomplish this, epigenetic mechanisms including RNA interference (RNAi), DNA methylation, and histone modifications are used to deal with the potentially dangerous effects of repeat transpositions and rearrangements (Slotkin and Martienssen, 2007; Maksakova et al., 2008).

Polycomb group proteins (PcG) are epigenetic repressors with the important function of maintaining the memory of transcriptional programs during development and differentiation (Morey and Helin, 2010; Schuettengruber et al., 2011). However, PcG role appears to go far beyond gene regulation, as they have been associated with many other important nuclear processes, including the regulation of higher order genome architecture and structure (Bantignies and Cavalli, 2011). Importantly, the vast majority of mammalian PcG proteins bind to non-coding DNA, and in particular repetitive elements, which for their intrinsic feature of being present in several copy number, may constitute binding platforms for Polycomb binding in mammals (Cabianca et al., 2012).

In this review, the biological role of DNA repeats and their epigenetic regulation is summarized with the hope of fostering new investigations of this largely unexplored region of the human genome.

GENETICS AND EPIGENETICS OF REPETITIVE ELEMENTS

Using classical annotation processes, about 50% of a typical mammalian genome is annotated as DNA repeats, 5–10% as genes and functional elements and the remaining 40–45% as DNA of unknown function. One caveat with traditional repeat annotation is that DNA repeat identification approaches, e.g., the RepeatMasker program (Smit et al., 1996–2004), use well-curated libraries of known repeat family consensus sequences. By doing so, ancient or divergent DNA repeat classes fail to be identified as repeats. Recently, using a highly sensitive alternative strategy, it was predicted that there may be more than 840 Mbp of additional repetitive sequences in the human genome, thus suggesting that up to 70% of the total genome is composed of repeats (De Koning et al., 2011).

DNA repeats can be present in different arrangements and sizes: they can be widely interspersed repeats (Table 1) or they can be located one next to another to form tandem repeats (Table 2). Repeats can range in size from 1 to 2 bases to millions of bases

and might comprise just two copies or millions of copies (Batzer and Deininger, 2002; Jurka et al., 2007; Kim et al., 2008; Britten, 2010; Hua-Van et al., 2011).

INTERSPERSED REPEATS

Interspersed repeats are the results of ancient or present activity of mobile genetic elements. These elements can mediate their own mobilization either by a cut-and-paste mechanism, as DNA transposons, or by a copy-and-paste process, like retro-transposons (Solyom and Kazazian, 2012). While DNA transposons are now considered immobile, some retro-transposon elements are able to mobilize themselves and other elements. Retrotransposons are composed of long terminal repeat (LTR) and non-LTR containing elements. The LTR retrotransposons are endogenous retroviruses (ERVs) that have lost the ability to go outside the host cell due to a non-functional envelope gene. Non-LTR retrotransposons can be subdivided into long interspersed elements (LINEs), short interspersed elements (SINEs) and, in hominid genomes, medium sized SVAs (SINE-R/VNTR (variable number of tandem repeat)/Alu). In humans, the most important LINE is the RNA polymerase II transcribed LINE-1 (L1), while SINEs are essentially represented by the RNA polymerase III transcribed Alus. L1 is the only element able to encode the proteins required for mobilization. Hence, these are the only known autonomously active human retrotransposons. L1s are also responsible for the mobilization of the non-autonomous Alus, SVAs and processed pseudogenes (cellular mRNAs that become substrates of the reverse transcriptases and are inserted into the genome).

Mobile elements have a significant role in evolution and in generating genetic diversity. For example, the genome fraction occupied by mobile elements varies in different species and each eukaryote displays a specific mobile element complement, suggesting that mobile elements are important players during speciation and evolution (Faulkner, 2011). Being significant contributors to the copy number variation present in humans, mobile elements are also an important source of genetic variation

Table 1 | Major features of the most represented interspersed repetitive elements in the human genome.

Repeat type				Estimated number of copies	Average length	Mobility	Estimated % genome coverage	
Interspersed	Retrotransposons	LTR	LTR (Long terminal repeat) or ERV (Endogenous retroviruses) (MaLR, ERV, ERV1, ERV-K, ERV-L, etc.)	200,000	6–11 kb	Autonomous retrotransposition (retroviral-like)	8%	42%
		Non-LTR	LINE (Long interspersed element) (L1, L2, CR1, etc.)	500,000	6 kb	Autonomous retrotransposition	20%	
			SINE (Short interspersed element) (Alu, MIR, etc.)	1,000,000	0.3 kb	L-1 dependent Retrotransposition	13%	
			SVA SINE-R/VNTR/Alu	2700	2–5 kb	L-1 dependent Retrotransposition		
	DNA transposons	DNA transposons (MER1, MER2, Mariner, Merlin, etc.)		300,000	1–3 kb	inert	2–3%	

Number of copies and genome coverage are estimated values based on current genome coverage.

Table 2 | Major features of the most represented tandem repeats in the human genome.

Repeat type			Unit length	Array length	Estimated % genome coverage
Tandem	Satellite	Alpha-Satellite	171 bp	3–5 Mb	22–25%
		Satellite II (HsatII)	23–26 bp or multiple	10–70 kb	
		Satellite III (GAATGn- simple sequence)	5 bp or multiples up to 70 bp	7.5–100 kb	
		Beta-Satellite	68 bp	2–14.5 kb	
		Gamma-Satellite	220 bp	10–200 kb	
	VNTR (Variable number of tandem repeats)	Microsatellite (Short tandem repeat)	1–13 bp	Hundreds bp	
		Minisatellite (including telomeric repeats)	6–100 bp	1–15 kb or more	
		Macrosatellite	2–12 kb or more	Tens up to hundreds bp	

Genome coverage is estimated on the basis of current genome coverage.

(Brouha et al., 2003; Bennett et al., 2004; Mills et al., 2007; Iskow et al., 2010; Ekram et al., 2012). Moreover, mobile elements can display differential activity in different tissues of the soma, suggesting that every individual is a genetic mosaic variegated by the differential insertion of mobile elements (Muotri et al., 2005, 2010). Finally, retrotransposons have recently been identified as a major source of epigenetic variations in the mammalian genome (Ekram et al., 2012). Retrotransposition, with only few exceptions such as V(D)J recombination (Brack et al., 1978), is an almost unique source of somatic genetic mosaicism, leading not only to heritable genetic variation but also to intra-individual variability. This represents a revolutionary concept that is changing the view of this class of repetitive elements (Faulkner, 2011).

Due to their nature, mobile elements have the potential to affect common diseases, through structural variation, deregulated transcriptional activity or epigenetic effects. Moreover, their transposition can directly cause insertional mutagenesis, as proved by the existence of nearly 100 examples of mobile element insertions causing disease (Lee et al., 2012; Solyom and Kazazian, 2012).

For their genotoxic potential, mobile elements are usually kept repressed by epigenetic mechanisms. DNA methylation represents one of the major players in the repression of repetitive elements (Liang et al., 2002; Kato et al., 2007). A recent study performed a comprehensive genome-wide methylation analysis on all repetitive elements in human embryonic stem cells and fetal fibroblasts (Su et al., 2012). Among all classes of repetitive elements, LINE, LTR, DNA transposon, and also satellite tandem DNA repeats appear more susceptible to changes in DNA methylation, thus suggesting that they are specifically regulated and silenced during cellular differentiation. Importantly, transposon-free regions (TFRs) in the genome have been selectively conserved and are associated with regions including CpG islands, suggesting that in mammalian genomes there are fragments of DNA that are largely unable to tolerate transposon insertion (Simons et al., 2006).

Aberrant repetitive DNA methylation can be associated with diseases. For example, hypo-methylation of L1, Alu, LTR, but also

of satellite repeats, is significantly associated with tumor progression in multiple cancers such as gastrointestinal stromal tumors, myeloma, and lung cancer (Rauch et al., 2008; Bollati et al., 2009; Igarashi et al., 2010).

Additionally, mobilization of L1 repeats has been associated with both physiological and pathological processes and is regulated by DNA methylation (Muotri et al., 2010). L1 mobilization has been associated with brain cell development, where the occurrence of L1 retrotransposition in adult cells has been suggested to contribute to neuronal somatic diversification (Muotri et al., 2005). This mechanism, so far assigned specifically to human neural progenitors and adult hippocampus, is modulated by the methyl-CpG-binding protein 2 (MeCP2) (Muotri et al., 2005, 2010; Coufal et al., 2009). Importantly, in RETT syndrome, a mental retardation disorder caused by mutation in the *MECP2* gene, an extensive de-regulation of L1 retrotransposition in neurons has been reported (Muotri et al., 2010; Solyom and Kazazian, 2012).

Besides DNA methylation, several repressive histone modifications, including H3K9me3, H3K27me3, and H4K20me3, are also enriched on interspersed repeats (Martens et al., 2005; Mikkelsen et al., 2007; Leeb et al., 2010). Importantly, a re-estimation of chromatin immunoprecipitation results on repetitive elements from high-throughput sequence data of human and mouse cells has been recently conducted (Day et al., 2010). According to this analysis, different members of the murine ERV family of repeats appear to assume distinct patterns of histone modifications, which are representative of a specific pattern of heterochromatin formation. While transposable elements belonging to ERV-K and ERV1 subfamilies are enriched for histone marks typical of constitutive heterochromatin such as H3K9me3 and H4K20me3 in mouse ES cells, ERV-L and MaLR families are characterized by the hallmark of Polycomb-mediated silencing H3K27me3 (Mikkelsen et al., 2007; Dong et al., 2008; Day et al., 2010).

A remarkable finding from these studies is that silencing of repetitive elements can be redundant and flexible. This has been shown by independent groups and within independent

silencing pathways. For example, during the stages of global DNA de-methylation in early embryonic mouse development, the RNA-interference guardian machinery become responsible for controlling the expression of intracisternal A particle (IAP), ERV-K, and ERV-L retrotransposons, thus preserving genome integrity (Svoboda et al., 2004). Additionally, studies of mouse ES cells deficient for the H3K9 histone methyltransferases Suv39h showed that decreased H3K9me3 levels in the repetitive elements were compensated by increases in H3K27me3 enrichment (Peters et al., 2003). Thus, different and largely independent repression pathways can converge and compensate each other's function. Most likely, this has to do with the necessity of the cells to guarantee multiple levels of protection from aberrant activation of mobile elements.

Overall, the epigenetic repression of repetitive elements on one hand prevents dramatic nuclear effects such as genotoxicity, but on the other hand allows the specific regulation of such elements occurring in the germ line (Peaston et al., 2004), embryonic cells (Kano et al., 2009) and, perhaps to a lesser extent, during later developmental phases (Muotri et al., 2005, 2010).

TANDEM REPEATS

Tandem repeats constitute a large portion of the human genome, and account for a significant amount of its copy number variation (Warburton et al., 2008). Besides their role in evolution (Warburton et al., 1996; Rudd et al., 2006; McLaughlin and Chadwick, 2011), they have been found to be critical in several other processes, including heterochromatin formation, chromosome segregation, (Morris and Moazed, 2007) and X-chromosome inactivation (XCI) (Chadwick, 2008). Moreover, repeat instability is at the basis of a number of diseases (Lopez Castel et al., 2010).

Tandem DNA repeats in the human genome show a wide range of unit sizes, spanning from a few base pairs in microsatellites, to several kilobases in megasatellites (Gelfand et al., 2007; Ames et al., 2008; Warburton et al., 2008). At a given locus, the tandem repeat copy number is usually polymorphic among individuals, and for this reason they are more commonly known as variable number tandem repeats (VNTRs).

One of the principal families of DNA tandem repeats in the genome is represented by the satellite DNA of chromosome centromeres. Indeed, maintenance of the structural integrity of centromeres and telomeres is one of the most important functions of tandem repeats (Blackburn, 1984). Centromeres have the fundamental role to ensure proper chromosome segregation during cell division. In the human genome, they consist of several Mb of alpha-satellite DNA, which is composed of a 171 bp repeat unit. Chromosome-specific higher-order repeat structures are typical of this type of repeat, as they are important for centromere function (Schueler et al., 2001). Forms of higher-order organization have also unexpectedly been characterized in "simple satellite" sequences such as GAATGn and VNTRs (Warburton et al., 2008), but whether this bears functional relevance has yet to be determined.

For their function, centromeres of higher eukaryotes require an epigenetic specification, rather than a defined DNA sequence. Indeed, centromeric regions localize in the pericentric

heterochromatic domain of the interphase nucleus, and they are enriched in H3K9me3, H4K20me3, H3K27me1 histone marks (Peters et al., 2001, 2003; Guenatri et al., 2004; Martens et al., 2005; Mikkelsen et al., 2007; Dong et al., 2008) and in proteins like the centromere-specific H3 variant Centromere protein A (CENP-A) (Yoda et al., 2000; Lo et al., 2001; Blower et al., 2002). As already described for the epigenetic regulation of interspersed repeats, loss of the H3K9 histone methyltransferases (HMTases) Suv39h, which are responsible for the tri-methylation of H3K9 (Peters et al., 2003), activates a compensatory mechanism leading to increase in H3K27me3 (a hallmark of Polycomb-mediated silencing). This underscores an unexpected plasticity between the H3K9 and H3K27 methylation systems (Peters et al., 2003).

In mice, where two different types of repetitive DNA sequences are associated with centromeres, major satellite repeats (6 megabases of 234 bp units) in the pericentromeric region, and minor satellite repeats (600 kb of 120 bp units) in the centromeric region (Choo, 1997), two distinct heterochromatic domains are distinguishable, which became important signatures of mouse interphase nuclei (Guenatri et al., 2004). Pericentromeric satellite DNA of different chromosomes forms large heterochromatic clusters, which upon DAPI staining result in DAPI-dense structures called chromocenters. These formations are typically enriched for the heterochromatin protein 1 alpha (HP1 α). The minor satellite DNA, instead, forms individual heterochromatin structures containing the CENP proteins (Guenatri et al., 2004).

In the human genome, the main groups of tandem repeats are the micro-, mini- or macro-satellites (Warburton et al., 2008). They are highly polymorphic in the general population and for this reason they are widely used as genetic markers. Macrosatellites consist of arrays of 1–12 kb repeat units, with a number of repeats ranging from a few to over one hundred (Warburton et al., 2008; Moseley et al., 2012). They can be either chromosome specific, as DXZ4 at chromosome Xq23 (Giacalone et al., 1992) and ZAV at chromosome 9q32 (Tremblay et al., 2010) or they can be associated with two or more chromosomal locations, such as D4Z4, on chromosomes 4q35 and 10q26; (Deidda et al., 1995; Winokur et al., 1996) and RS447, on 4p15 and 18p23; (Gondo et al., 1998).

DXZ4 and D4Z4 macrosatellites are both extensively regulated at the epigenetic level, and they have been described as being associated with either euchromatic or heterochromatic states. Contraction of the 3.3 kb polymorphic D4Z4 tandem repeat array on chromosome 4q35 is associated with facioscapulohumeral muscular dystrophy (FSHD) where a shortening below the threshold of 11 repeat units generates an epigenetic and topologic remodeling of the locus, thus leading to the pathology (Cabanca and Gabellini, 2010). The X-linked DXZ4 macrosatellite locus, instead, has an opposing conformation to that of the surrounding chromosome, constituting a euchromatic dot in the inactive X chromosome, and vice versa (Chadwick, 2008). For their very peculiar epigenetic features and for their involvement in fundamental biological and pathological processes, D4Z4 and DXZ4 could emerge as paradigms for understanding the epigenetic regulation of tandem DNA.

D4Z4 AND DXZ4

Two of the most extensively investigated macrosatellites are the X-linked DXZ4 and the chromosome 4-linked D4Z4. Despite lacking sequence similarity, D4Z4 and DXZ4 macrosatellites share several common aspects (Chadwick, 2009). DXZ4 and D4Z4 are extremely GC rich and belong to a family of human macrosatellites that are noncentromerically located (Giacalone et al., 1992; Kogi et al., 1997; Chadwick, 2009; Tremblay et al., 2010).

Each DXZ4 unit is 3.0 kb long and organized in a tandem array containing 12 to more than 100 copies, localized at Xq23 (Giacalone et al., 1992). As typical for an X-linked locus, DXZ4 is hemizygous in males and subject to XCI in females. However, DXZ4 adopts an opposite chromatin conformation compared to that of the surrounding X chromosome. In males and on the active X-chromosome (Xa), DXZ4 displays features of constitutive heterochromatin, like enrichments in the repressive histone mark H3K9me3, high levels of DNA methylation and association with heterochromatin protein 1 gamma (HP1 γ). On the contrary, in the inactive X (Xi), DXZ4 is characterized by euchromatic histone marks such as H3K4me2 and H3K9Ac, a low level of DNA methylation, and is bound by the chromatin regulators CTCF and YY1 (Chadwick, 2008; Filippova, 2008; Moseley et al., 2012). Notably, these features of DXZ4 are remarkably similar to those of the mouse X-inactivation center (Xic), a region of the X chromosome required for XCI (Courtier et al., 1995; Chao et al., 2002; Boumil et al., 2006; Donohoe et al., 2007). Finally, DXZ4 resides at the distal edge of a heterochromatic region targeted by PcG epigenetic repressors (Chadwick and Willard, 2004; McLaughlin and Chadwick, 2011).

The D4Z4 macrosatellite maps to the subtelomeric region of the chromosome 4 long arm, in 4q35. Each unit is 3.3 kb and is present in 11 to 100–150 copies in the general population. Interestingly, reduction of D4Z4 copy number below 11 units is associated with FSHD, one of the most important forms of muscular dystrophy (Wijmenga et al., 1992; Van Deutekom et al., 1993). D4Z4 belongs to a family of repeats with high sequence identity present also in human chromosomes 10q26, 1p12, and the p-arm of acrocentric chromosomes (Lyle et al., 1995; Winokur et al., 1996). This results in frequent exchanges between the 4q35 and 10q26 arrays, which share the highest identity (Van Deutekom et al., 1993). Like DXZ4, D4Z4 is bound by the epigenetic factor YY1 (Gabellini et al., 2002) and displays alternative epigenetic states that parallel the ones of DXZ4 in Xa versus Xi. For D4Z4, the epigenetic make-up is copy number-dependent. The non-contracted array, which retains more than 11 D4Z4 units, displays heterochromatic features like the repressive histone marks H3K9me3 (Zeng et al., 2009) and H3K27me3 (Bodega et al., 2009; Cabianca et al., 2012), histone hypoacetylation (Jiang et al., 2003), as well as a high level of DNA methylation (Van Overveld et al., 2003). Reduction of D4Z4 copy number below 11 units is associated with reduced levels of repressive histone marks (Bodega et al., 2009; Zeng et al., 2009; Cabianca et al., 2012), acquisition of the activating histone marks H3K4me3 and H3K36me2 (Cabianca et al., 2012), DNA hypomethylation (Van Overveld et al., 2003), binding of CTCF (Ottaviani et al., 2009) and loss of Polycomb silencing (Cabianca et al., 2012).

Like DXZ4, D4Z4 is bi-directionally transcribed to generate non-protein-coding RNAs (ncRNAs) (Chadwick, 2008; Snider et al., 2009; Tremblay et al., 2011; Block et al., 2012; Cabianca et al., 2012). In particular, D4Z4 generates a long, chromatin-associated ncRNA (*DBE-T*) selectively in FSHD patients. *DBE-T* functions *in cis* by recruiting the Trithorax protein ASH1L to the FSHD locus leading to chromatin remodeling and de-repression of 4q35 genes (Cabianca et al., 2012). Hence, similarly to the dichotomous behavior observed for DXZ4 on Xi and Xa chromosomes, for D4Z4 the FSHD pathogenesis underlies a major epigenetic switch from a Polycomb repressed state to a Trithorax de-repressed state.

The last, most telomeric D4Z4 unit at 4q35 encodes for a protein called DUX4 (double homeobox 4), which represents one of the major candidates for FSHD (Lemmers et al., 2010). The *DUX4* gene itself originates from a repetitive element, as it is a processed pseudogene of the ancestral *DUXC* gene. Interestingly, *DUX4* and not *DUXC* has been selectively retained in the primate lineage (Clapp et al., 2007; Leidenroth and Hewitt, 2010). In healthy subjects *DUX4* is expressed only in the germ line, while it is epigenetically silenced in somatic tissues (Snider et al., 2010). In FSHD, *DUX4* is aberrantly expressed in skeletal muscle (Dixit et al., 2007; Snider et al., 2010).

DUX4 protein is a transcriptional activator able to bind and activate transcription of MaLR repetitive elements (Geng et al., 2012). Interestingly, MaLR retrotransposons are known Polycomb targets (Day et al., 2010). Hence, DUX4 could have the physiological role of collaborating with Polycomb for the regulation of repetitive elements during early developmental stages and in the germ line.

POLYCOMB

PcG proteins and their functional counterpart, the Trithorax Group (TrxG) proteins, are evolutionary-conserved chromatin regulatory factors that were originally identified in *Drosophila* (Schuettengruber et al., 2007, 2011; Morey and Helin, 2010). PcG and TrxG are essential for cellular identity and differentiation in multicellular organisms. Their activity is required to maintain an “epigenetic memory” of specific gene expression patterns. This is at the basis of the establishment of the correct spatio-temporal regulation of gene expression and, more importantly, of its transmission throughout cell division and cell fate choices. In general, PcG collaborates with transcriptional repressors to maintain gene silencing while TrxG works by counteracting PcG activity allowing, if the appropriate transcriptional activators are available, for gene activation (Schuettengruber et al., 2007). In vertebrates, PcG and TrxG play a central role in stem-cell plasticity and renewal, proliferation, genomic imprinting, X-inactivation, and cancer (Schuettengruber et al., 2007).

In *Drosophila*, where the Polycomb system was first described, PcG and TrxG are specifically recruited on so-called Polycomb Response Element (PRE)/Trithorax Response Element (TRE) sequences, which are switchable memory DNA modules, with PcG or TrxG as their effectors (Schuettengruber et al., 2011). The mechanisms underlying PcG recruitment in mammals are still controversial, though some vertebrate PRE-like elements have recently been described. Interestingly, these retain features of

Drosophila PREs including binding sites for DNA-binding of factors involved in PcG recruitment to PREs in *Drosophila* (Sing et al., 2009; Woo et al., 2010; Cuddapah et al., 2012). However, a defined role for mammalian homologs of PcG recruiters has not been established. Accordingly, additional mechanisms for PcG recruitment in mammals have been proposed. Several examples for a role of short and long ncRNAs in PcG recruitment in ES cells are available (Rinn et al., 2007; Zhao et al., 2008; Khalil et al., 2009; Gupta et al., 2010; Kanhere et al., 2010; Guil et al., 2012). Moreover, in mammals there is a strong correlation between PcG binding and CpG islands (Tanay et al., 2007; Ku et al., 2008; Mendenhall et al., 2010). In particular, non-methylated GC-rich sequences depleted of activating motifs have been shown to be sufficient for Polycomb recruitment in mammalian embryonic stem cells (Mendenhall et al., 2010).

Polycomb proteins form two major multiprotein complexes, Polycomb Repressive Complex 1 (PRC1) and 2 (PRC2). *Drosophila* PRC1 displays four core subunits: Polycomb (Pc), Polyhomeotic (Ph), Posterior sex combs (Psc), and Sex combs extra (Sce, also called dRing). PRC2 core subunits are Enhancer of zeste, E(z), Extra sex combs (Esc), Suppressor of zeste 12, Su(z)12, and the nucleosome-remodeling factor 55 (Nurf-55). In vertebrates, PRC1 and PRC2 are conserved in overall organization, but display a higher complexity in terms of subunits and interactions, so that their composition is cell type- and developmental stage-dependent (Kuzmichev et al., 2002, 2004, 2005; Gao et al., 2012).

Both PRC1 and PRC2 complexes retain an enzymatic activity. In PRC1, the RING domain containing protein dRing (Ring1B in vertebrates) is an E3 ubiquitin ligase mediating the ubiquitination of lysine 119 on histone H2A, which has been suggested to induce chromatin compaction and inhibit transcription elongation (De Napoles et al., 2004; Wang et al., 2004). Nevertheless, in the case of Ring1B the requirement of the enzymatic activity for chromatin compaction was recently challenged (Eskeland et al., 2010). PRC2 catalyzes the di-methylation and tri-methylation of histone H3 at lysine 27 (H3K27me2/me3). The catalytic subunit of PRC2, E(z) in flies, Enhancer of zeste homologs 1/2 (Ezh1/Ezh2) in vertebrates, contains the SET histone methyltransferase domain (Morey and Helin, 2010). Importantly, for its activity, E(z) requires the binding of two other PRC2 core components, Su(z)12/suppressor of zeste 12 (Suz12), and Esc/embryonic ectoderm development (Eed) (Morey and Helin, 2010).

H3K27me3 is a fundamental histone mark (hallmark) of Polycomb binding. Frequently, H3K27me3 is spread out to broad regions marking large PcG domains allowing for PRE-mediated repression several tens of kilobases away from target genes (Schuettengruber et al., 2007; Morey and Helin, 2010). H3K27me3 also represents a docking site recognized by PC (Cbx in vertebrates) contained in the PRC1 complex. Based on this, a sequential PRC2, PRC1 recruitment has been proposed (Cao et al., 2002). Nevertheless, it was recently shown that PRC1 recruitment to target genes in mammals can be also independent from PRC2 (Gao et al., 2012; Tavares et al., 2012).

POLYCOMB AND REPEATS

Polycomb-associated histone marks are prevalent in the mammalian genome. Quantitative mass spectrometry studies reported

that up to 70% of histone H3 carries the PRC2 histone marks H3K27me2 or me3 (Peters et al., 2003; Schoeftner et al., 2006). However, genes and known functional elements comprise only up to 10% of the mammalian genome (Pheasant and Mattick, 2007), while over two-thirds of the remaining part is composed of repetitive elements (De Koning et al., 2011). Hence, this simple observation raises interesting questions about the possible acquirement of novel functions by the PcG proteins along with evolution, involving the non-coding fraction of the mammalian DNA.

Several reports show the presence of Polycomb repressive histone marks on repetitive elements. Initially, PcG silencing on repeats was described as a compensatory mechanism upon loss of H3K9me3 repression in pericentric DNA, where H3K27me1 was converted into H3K27me3 (Peters et al., 2003). More recently, the characterization of the epigenetic pattern of ERV-L and MaLR retrotransposons revealed that they are marked by H3K27me3 (Day et al., 2010), and importantly, a crucial role for ERV-L retrotransposons in embryo totipotency and development has been described (Macfarlan et al., 2011, 2012). At the very early two-cell stage, the murine endogenous retroviral elements ERVL (MuERV-L) are transiently de-repressed (Kigami et al., 2003). Their expression is significant, as it represents 3% of the total transcriptional output, and it is very sharply regulated in time, as it is specific for the developmental stage of the embryo where blastomeres are still totipotent (Svoboda et al., 2004).

Importantly, ERVL transcripts represent a source of regulatory elements which is co-opted by cellular genes to co-regulate their cell stage-specific expression (Macfarlan et al., 2012). In this process, more than 25% of MuERV-L copies are activated and 307 protein-coding genes generate 626 different chimeric transcripts with MuERV-L elements. Among the genes that use alternative MuERV-L-LTR promoters to initiate their transcription, there are transcription factors like Gata-4, which is involved in lineage determination and embryo development (Soudais et al., 1995) and is a known PcG target (Tiwari et al., 2008). Remarkably, MuERV-L expression is regulated by histone modifications like H3K4me3, the active histone mark typical of TrxG proteins (Schuettengruber et al., 2011). In fact, in the absence of the H3K4me3 demethylase LSD1/KDM1A, which is critical for the H3K4/H3K27 methylation balance in human ES cells (Adamo et al., 2011), MuERV-L/MERVL becomes overexpressed and embryonic development arrests at gastrulation (Macfarlan et al., 2011). Given that Polycomb and Trithorax are the major players in development and ERV-L is repressed via PcG mediated-silencing (Day et al., 2010), it is tempting to speculate that MuERV-L retrotransposons undergo a Polycomb/Trithorax regulation, with Polycomb mediating their repression and Trithorax their spatiotemporal-specific up-regulation in order to drive cell-fate specification.

A direct link between Polycomb and repeats-mediated silencing has been recently reported (Leeb et al., 2010). This work, in fact, not only identified both murine leukemia virus (MLV) and IAP retroelements as targets of Polycomb complexes, but also performed the first PcG loss-of-function study in a genomic repeat contest. Indeed, upon double knock out of key PRC1 and PRC2 components, Leeb et al. observed a strong increase

in expression of LTR retrotransposons, which in turn provoked their active mobilization (Leeb et al., 2010). In particular, both MLV and IAP elements were found strongly de-repressed in ES cells double null for the Polycomb proteins Eed and Ring1B when compared to both wild type and single KO cells. Importantly, loss of binding of Polycomb complexes on MLV and the subsequent de-repression of these elements was associated with a considerable increase in MLV mobilization (Leeb et al., 2010). Similarly, Eed^{-/-} Ring1B^{-/-} ES cells showed IAP de-repression, that was associated with reduced levels of DNA methylation on IAP repeats in the double KO and Eed^{-/-} ES cells, in agreement with a previous report about repressive function on IAP retroelements of DNA methylation (Walsh et al., 1998). Hence, this work once again showed a redundancy in the mechanisms of repeat silencing, similarly to that previously reported for other repressive histone marks (Peters et al., 2003; Svoboda et al., 2004). Both PRC1 and PRC2 complexes, in fact, are recruited in parallel for LTR PcG-mediated silencing, as the single KO produced only a partial effect of de-repression (Leeb et al., 2010), thus suggesting that mechanisms of retrotransposon repression act redundantly even when mediated by the Polycomb machinery. Based on these results, it was suggested that genomic repeats, for their intrinsic feature of being present in several copies in the genome, could constitute binding platforms for mammalian PcG complexes (Leeb et al., 2010). Notably, epigenetic silencing of transgenes present in multiple copies has been already described in mice (Garrick et al., 1998; Festenstein et al., 1999; Hirasami and Festenstein, 2005) and it is well-established that proximity of DNA binding sites encourages cooperation among transcription factors (Amouyal et al., 1998; Amouyal, 2007).

Since the greatest proportion of Polycomb-mediated chromatin modifications is located in non-genic regions, a loss of PcG activity would need to be considered not only for its specific effect on Polycomb targets, but also for its possible effects on genome stability.

ncRNAs IN A POLYCOMB AND REPEAT LANDSCAPE

Repeats can be specifically transcribed. Around 6–30% of the total amount of transcripts in mammalian cells initiates within repetitive elements and their expression is frequently tissue-specific (Faulkner et al., 2009). Recent studies show that repeats play central roles in regulating gene expression at multiple levels (Norris et al., 1995; Speek, 2001; Faulkner and Carninci, 2009; Kaneko et al., 2011; Shen et al., 2011). Repetitive elements may regulate the expression of nearby protein-coding genes by providing tissue-specific promoters or enhancers (Speek, 2001; Conley et al., 2008; Faulkner et al., 2009); they can be co-opted to generate alternative exons (Zhang and Chasin, 2006); they can modulate the abundance of gene products, for example through generation of ncRNAs, working *in trans* or *in cis*, either enhancing (by anti-silencing) or reducing (by transcriptional interference) their expression (Allen et al., 2004); or they can produce short ncRNAs exploited by RNAi machinery (Ghildiyal et al., 2008; Watanabe et al., 2008; Faulkner and Carninci, 2009) (Figure 1).

If the regulatory functions are combined with the ability of retrotransposons to mobilize upon de-repression or of tandem repeats to rearrange during meiotic division, the scenario becomes even more complex, as novel insertions of mobile repetitive elements or change in tandem repeat copy number may

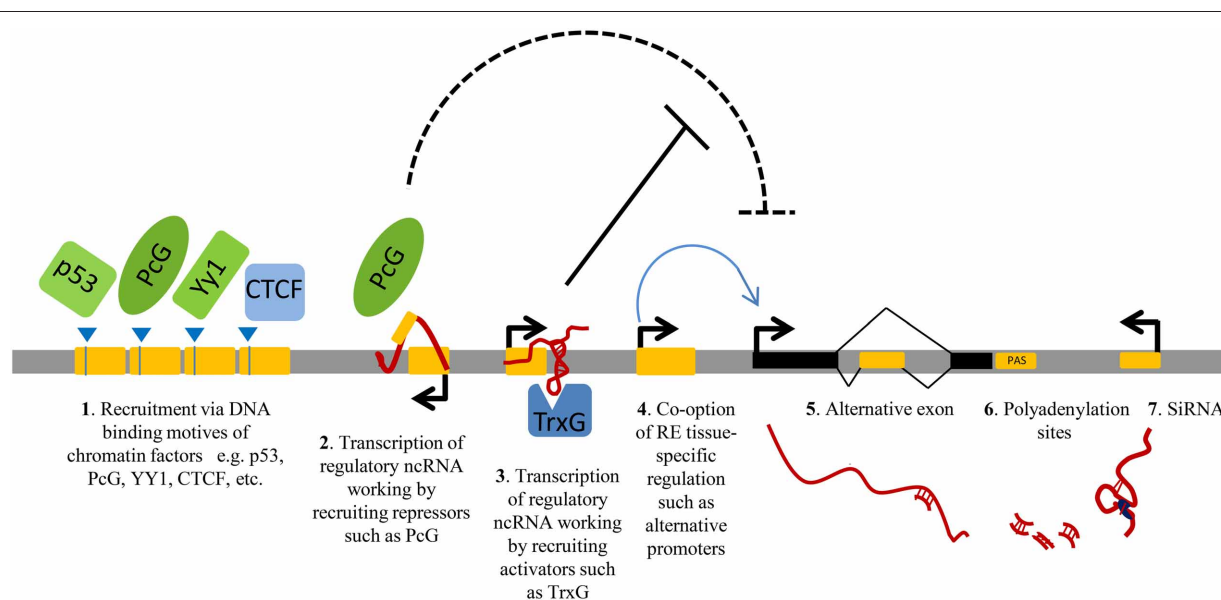


FIGURE 1 | A schematic view of the principal gene-regulatory functions of repeats. Repetitive elements (RE, represented as yellow modules) have an impact on gene expression by providing DNA binding sites for transcription factors or chromatin regulators (1); upon transcription, by generating regulatory non-protein-coding RNAs (ncRNAs) involved in gene silencing (2) or gene activation (3), for example via

direct recruitment of repressors/activators such as Polycomb (PcG) and Trithorax (TrxG) (2–3). Repeat sequences can also contribute to gene transcription by providing alternative promoters (4); alternative exons (5); polyadenylation sites (PAS, 6) or they can influence the stability of other transcripts via RNA interference (RNAi) by producing short double-stranded RNAs (dsRNAs) (7).

modify the chromatin structure (Lunyak et al., 2007) and the gene regulation of nearby genes (Cabianca et al., 2012).

Important examples of the interplay between Polycomb, repeats and ncRNAs in normal physiology and in disease are illustrated in the following sections.

X-INACTIVATION

X-inactivation, the process that leads to the silencing of one X chromosome in mammalian female cells, represents one of the most striking examples of long-range chromosomal regulation involving ncRNAs, Polycomb-mediating silencing and DNA repeats (Hall and Lawrence, 2010). In mammals, a large non-coding RNA named *Xist* “paints” the X-chromosome *in cis* (Brockdorff et al., 1992; Brown et al., 1992; Clemson et al., 2006; Chow et al., 2007) and induces a silencing cascade repressing the whole chromosome territory (Hall and Lawrence, 2003; Heard and Disteche, 2006). *Xist* works by recruiting PRC1, PRC2 and their respective histone marks (Leeb and Wutz, 2007) to the core of the inactive X chromosome, which contains genomic repeats (Chaumeil et al., 2006; Clemson et al., 2006). Besides local changes, a higher-order remodeling of the chromatin architecture takes place, thus producing the well-known silent core corresponding to the DAPI-dense Barr Body, which resides in the heterochromatic compartment at the nuclear or nucleolar periphery (Clemson et al., 2006).

Different classes of repeats play their roles in X-inactivation. Common repeats, like LINE-1 and Alu, participate structurally in the formation of the heterochromatic inner core of the Xi DNA territory (Hall and Lawrence, 2010), whereas a role for the euchromatic DXZ4 macrosatellite locus in Xi chromosome has been suggested (Chadwick, 2008). Moreover, the *Xist* ncRNA contains several tandem repeats termed A, B, C, D, E, and F (Hendrich et al., 1997; Nesterova et al., 2001; Yen et al., 2007; Horvath et al., 2011). Repeat A, with its conserved sequence and tetra-loop structure (Duszczek et al., 2011), is essential for Polycomb-mediated silencing of X-linked genes (Wutz et al., 2002; Zhao et al., 2008). In fact, in the future Xi chromosome, PRC2 is initially recruited by the 1.6 kb *RepA* ncRNA, which is directly bound by the PRC2 subunit Ezh2. The *RepA*/PRC2 interaction enables the full-length *Xist* induction and thus the spreading of the *Xist* ncRNA and PcG silencing on the whole Xi chromosome (Zhao et al., 2008). The *RepA* region is the primary target of PcG binding also within the 17 kb full-length ncRNA *Xist* (Zhao et al., 2008), and indeed in *RepA* mutants, *Xist* recruits 80–90% less PRC2 (Kohlmaier et al., 2004).

The antisense 40 kb *Tsix* ncRNA is able to inhibit the *RepA*/Ezh2 interaction, probably by competing with *Xist* for PRC2 binding (Zhao et al., 2008). In pre-XCI cells, *Tsix* keeps in check the state of both X chromosomes and only a few molecules of *Xist* are transcribed (Zhao et al., 2008). When cell differentiation triggers dosage compensation, another regulatory ncRNA named *Jpx* becomes actively transcribed from the *Xist* loci of both X chromosomes, thus supplying the required activator for high-level *Xist* expression (Tian et al., 2010). In the future Xi, *Tsix* is now down-regulated, hence producing a permissive state for *Xist* induction, whereas, in the future Xa, the levels of *Tsix*

continue titrating away PcG from *RepA*, thus maintaining blocked the repressive cascade (Zhao et al., 2008).

Another important *Xist* repeat is Repeat C, a C-rich sequence, specific of *Xist* and highly conserved, which is important for *Xist* localization on the inactive X chromosome (Memili et al., 2001; Sarma et al., 2010). A recent report provided an important role for another repeat of the *Xist* locus, Repeat F (Jeon and Lee, 2011). This region, characterized by the presence of CTCF and YY1 binding sites, is bound by YY1, which with its multiple zinc fingers is able to bind both DNA and RNA at the same time. YY1 bridges the *Xist* ncRNA via Repeat C (Sarma et al., 2010), and the X chromosome, via the Repeat F region. Overall, X inactivation provides a strong argument for an important physiological interplay between repeats, Polycomb, and ncRNAs (Figure 2).

FSHD MUSCULAR DYSTROPHY

FSHD (OMIM 158900) is a genetic disorder of particular interest for the atypical interactions between genetic and epigenetic players, which both contribute to the etiology of the disease (Neguembor and Gabellini, 2010). FSHD is an autosomal dominant disease and for more than 20 years it has been known to be associated with reduction in copy number of a macrosatellite repeat (called D4Z4) mapping to the subtelomeric 4q35 region (Wijmenga et al., 1990, 1991, 1992; Van Deutekom et al., 1993). Also, it has been known for a decade that D4Z4 deletions cause de-repression of genes located nearby (Gabellini et al., 2002). Nevertheless, the molecular understanding of the D4Z4 repeat mechanism of action was only recently provided (Cabianca et al., 2012).

Each D4Z4 unit is extremely GC rich, containing a sequence nearly identical to the consensus motif of *Drosophila* PREs and several putative DNA binding sites for factors which are Polycomb recruiters in *Drosophila*, such as YY1 and GAGA factor (Mihaly et al., 1998; Busturia et al., 2001; Mishra et al., 2001; Gabellini et al., 2002; Cabianca et al., 2012). Accordingly, in healthy subjects the D4Z4 tandem array is extensively bound by PRC1 and PRC2 and displays enrichment for the typical PcG-associated repressive histone marks H2AK119Ub and H3K27me3. The region is also bound by proteins associated to Polycomb recruitment in mammals like Jarid2 (Peng et al., 2009; Shen et al., 2009; Landeira et al., 2010; Li et al., 2010; Pasini et al., 2010) or homologs of PcG recruiters in *Drosophila* (YY1, HMGB2, c-Krox/Th-POK; vertebrate fly homologs Pho, Dsp1, GAGA factor, respectively) (Busturia et al., 2001; Mishra et al., 2001; Gabellini et al., 2002; Dejardin et al., 2005; Matharu et al., 2010). Finally, the repeats array also shows enrichment for the Polycomb-associated histone variant macroH2A (Buschbeck et al., 2009).

Importantly, D4Z4 is able to initiate PcG recruitment to ectopic sites and mediate copy number-dependent repression of gene expression, typical features of *Drosophila* PREs (Gabellini et al., 2002; Cabianca et al., 2012). In FSHD patients, the reduction in D4Z4 copy number is associated with a reduction in PcG silencing. This allows for the production of a long, chromatin-associated ncRNA: *DBE-T*. *DBE-T* works *in cis* by directly recruiting the TrxG protein ASH1L to the 4q35 locus. This leads to a structural and epigenetic remodeling of the FSHD locus, toward a more active chromatin state, which is responsible

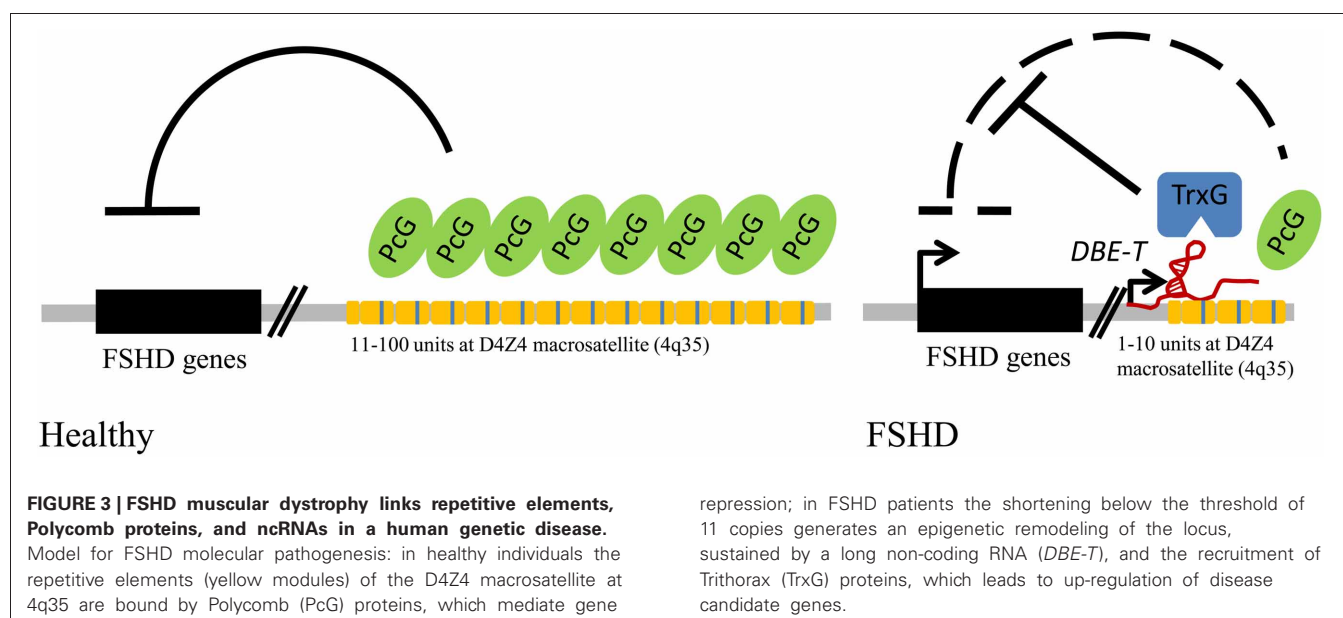
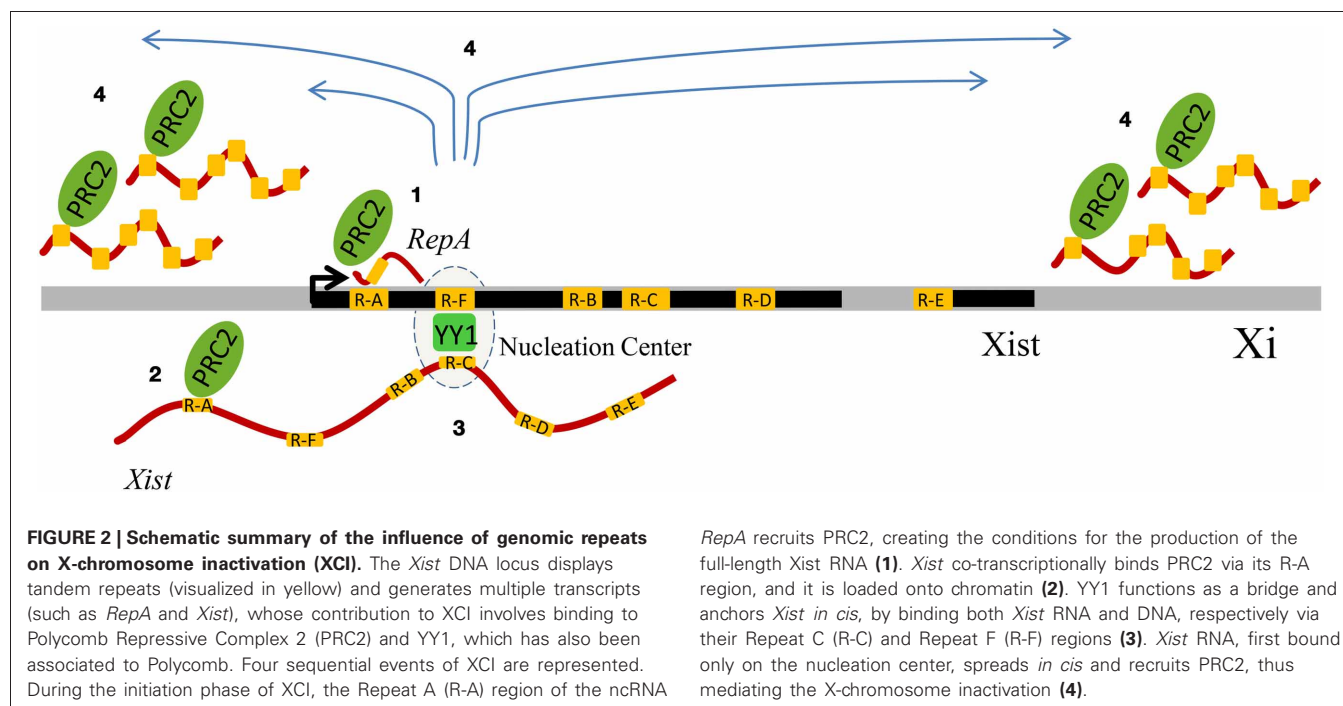
for the de-repression of 4q35 genes. Altogether, FSHD constitutes an important example of the relevance of DNA repeats, Polycomb and ncRNAs in human genetic diseases (Figure 3).

REPETITIVE ELEMENTS AND CHROMATIN ORGANIZATION IN THE 3D NUCLEAR SPACE

In general, nuclear organization of chromatin reflects its active or inactive state. Euchromatin occupies the internal nucleoplasm, whereas heterochromatin preferentially localizes at the nuclear and nucleolar periphery (Kosak et al., 2002; Shopland et al., 2003). Accordingly, repetitive elements can also localize

differently. For example, pericentromeric satellite repeats are usually confined to the heterochromatic domains of the nuclear periphery whereas telomeres of human chromosomes usually reside in the internal compartment (Tam et al., 2004). There are important exceptions; the FSHD-associated 4q35 telomere behaves differentially, being usually associated to the nuclear periphery (Masny et al., 2004; Tam et al., 2004).

The nuclear machineries are not uniformly distributed in the nucleoplasm, but are organized in functional sub-compartments, so-called “factories” or “hubs” (Lamond and Spector, 2003; Hall et al., 2006; Meaburn and Misteli, 2007). In fact, by staining for



a particular key factor of important nuclear processes (like transcription, RNA processing, replication, or DNA repair), a number of discrete structures appear in the nucleus, which result from the local concentration of proteins involved in specific nuclear processes. For example, “transcription factories” have been described and different genes, localized on distant chromosomal loci, can associate to the same active foci to be co-transcribed (Osborne et al., 2004). For nuclear compartments, patterns of distribution in the nucleus, characteristic of the different cell type or differentiation state, can be recognized (Lancot et al., 2007). However, it is still an open question whether a fragment of DNA needs to be primarily attracted to one of these nuclear compartments in order to be functionally processed, or if the specific machinery can also activate elsewhere in the nucleus but needs to reach these structures for a higher efficiency.

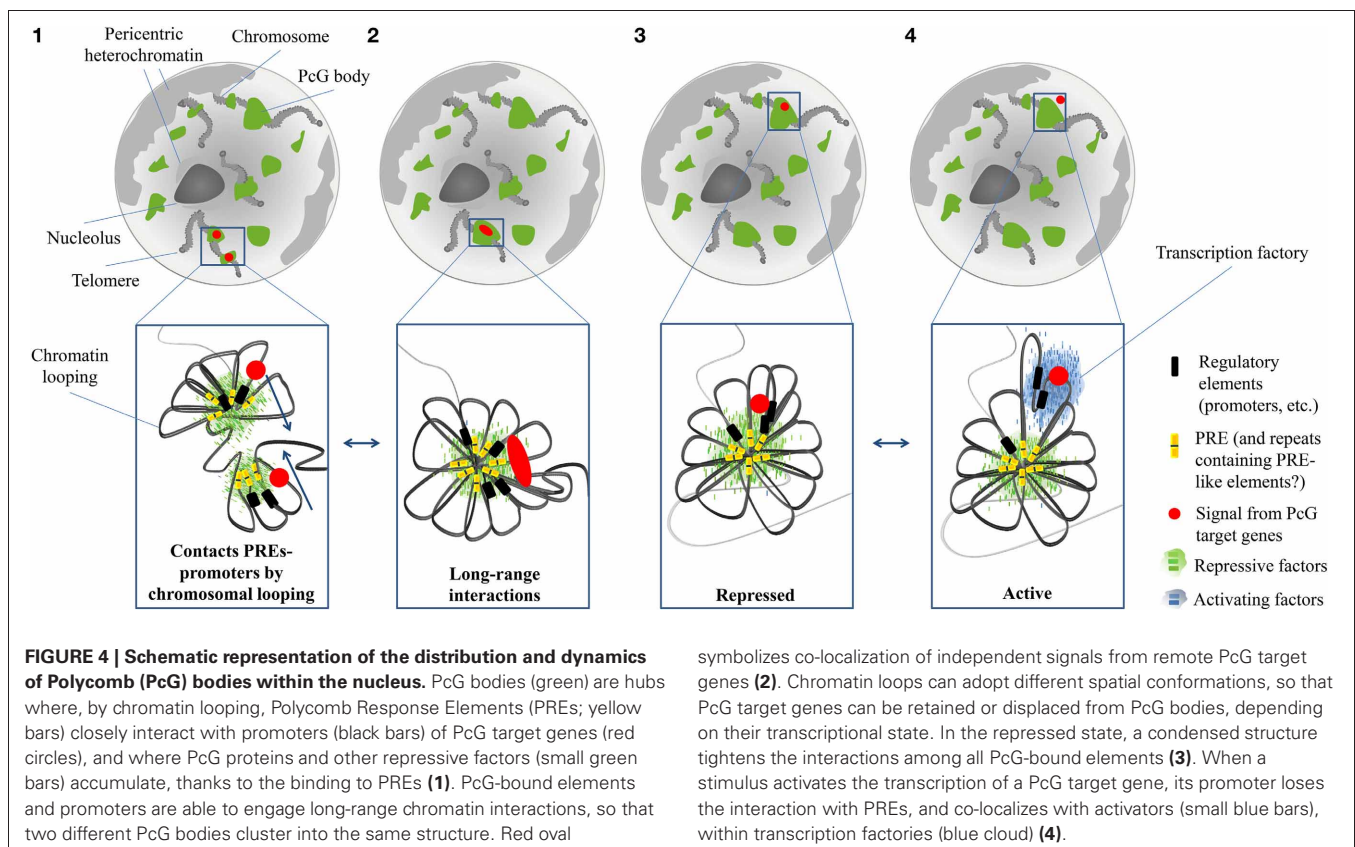
Polycomb proteins and associated histone marks reside in discrete nuclear structures called Polycomb bodies, co-localizing with stably repressed homeotic genes (Messmer et al., 1992; Buchenau et al., 1998; Grimaud et al., 2006; Ferraiuolo et al., 2010; Bantignies et al., 2011). These repressive chromatin hubs are composed of chromatin loops involving PcG-bound regulatory elements and promoters of PcG target genes (Cleard et al., 2006; Comet et al., 2011). Hence not only events associated with gene activation, but also those associated with gene repression, including the ones involving Polycomb proteins, can localize on discrete foci, where long-range interactions take place.

The organization of these structures in *Drosophila* starts at the level of PREs, the DNA modules recruiting Polycomb complexes

(Muller and Kassis, 2006; Schuettengruber et al., 2007). As already discussed, the histone-methylation activity of the PRC2 complex spreads out on neighboring regions, marking large PcG domains. Hence, PcG silencing reaches target genes that are tens of kilobases distant from a PRE. Moreover, PREs tend to cluster in larger domains (Bantignies and Cavalli, 2011) (Figure 4).

As characterized by chromosome conformation capture experiments, long distance intra- and even inter-chromosomal interactions among PcG targets are established, thus producing a major level of chromatin organization in the 3D nuclear space (Lanzuolo et al., 2007; Terranova et al., 2008; Tiwari et al., 2008; Eskeland et al., 2010; Comet et al., 2011; Tolhuis et al., 2011). It has been proposed that these long-range contacts are mediated by ncRNAs (Rinn et al., 2007), insulators DNA element (Li et al., 2011) and RNAi machinery (Grimaud et al., 2006). On top of such a hierarchical organization of PcG domains are found the PcG bodies. PcG bodies differ in size and Polycomb intensity. In particular, PcG domains with a larger linear size display a higher content of Polycomb and generate bigger and more intense PcG bodies (Cheutin and Cavalli, 2012).

The discovery of PcG bodies raised questions about their function: are they merely the result of the accumulation of PcG proteins to clustered Polycomb domains, or is the formation of these “hubs” required for PcG silencing (Buchenau et al., 1998)? The fact that PcG proteins organize in such PcG bodies instead of being uniformly distributed in the nucleus is already an indication toward a functional role for these structures. Indeed, PcG-mediated gene silencing occurs within PcG bodies



(Grimaud et al., 2006) and it has been proposed that the local concentration of PcG components and their target genes in PcG bodies may produce chromatin condensation (Terranova et al., 2008; Eskeland et al., 2010). Indeed, a correlation between repression of PcG targets and their localization in PcG bodies has been reported. For example, Fab-7, the PRE-containing region controlling the expression of the gene *Abd-B*, is found within PcG bodies when *Abd-B* is repressed, whereas it is outside the PcG bodies when *Abd-B* is expressed (Lanzuolo et al., 2007; Bantignies and Cavalli, 2011; Bantignies et al., 2011). Active genes are displaced from these repressive chromatin hubs not only in *Drosophila* but also in mammals. For example, the human GATA-4 locus, involving several PcG bound regions, shows a similar chromatin structure depending on its transcriptional state (Tiware et al., 2008) (Figure 4).

As it primarily functions as a marker regulator of development, Polycomb accumulation, and thus the presence of PcG bodies are regulated during cell differentiation. Experiments of fluorescence recovery after photobleaching (FRAP) in both *Drosophila* and mammalian embryonic stem cells, showed a dynamic exchange of PcG proteins between PcG bodies and nucleoplasm (Ficz et al., 2005; Ren et al., 2008). In *Drosophila*, Polycomb starts accumulating in the nucleus during the early stages of development (stage 5), progressively increases and gets recruited to PcG bodies (stages 5–11), until it becomes stably associated with PcG bodies during late embryogenesis (Cheutin and Cavalli, 2012). To address the question of whether the formation of PcG bodies is the direct result of PcG binding to their targets or, on the contrary, PcG targets need to associate with PcG bodies in order to be repressed, *in vivo* live imaging approaches have been used to characterize the motion of PcG targets and PcG bodies in the nucleus. Interestingly, a motion away from PcG bodies from the nuclear periphery toward the nuclear interior, regulated by actin and nuclear myosin I, was observed immediately after inducing transcription (Chuang et al., 2006). Similarly to other chromatin domains, Polycomb bodies' motion sensitively decreases upon differentiation, and shows similar kinetics, either fast but limited to volumes much smaller than chromosome territory occupancy, or slow but involving overall a higher level of nuclear structure (Cheutin and Cavalli, 2012).

Based on the fact that Polycomb is concentrated in PcG bodies by immunofluorescence and in repeats by chromatin immunoprecipitation, it could be hypothesized that genomic repeats which are Polycomb targets in mammals could functionally behave in a similar way to PREs and mediate association between Polycomb-regulated genes. In this view, Polycomb complexes and repetitive elements would play a role in the compartmentalization of the nucleus, establishing large chromatin domains where PcG target genes are efficiently repressed. Interestingly, it has been shown that the 3D organization of PcG target genes can influence PcG-mediated silencing. In *Drosophila* the deletion of Fab-7 perturbed the interaction between BX-C and ANT-C, producing mild effects on gene expression at distant Polycomb target genes. However, sensitized genetic backgrounds had to be used in order to observe homeotic phenotypes (Bantignies et al., 2011). Interestingly, in mammals structural alterations of repetitive sequences can affect long-range PcG-mediated silencing *in cis* (Cabanca et al., 2012). Moreover, deletions or mutations of genetic elements on one chromosome can affect expression of interacting genes *in trans* (Spilianakis et al., 2005; Ling et al., 2006).

Collectively, these considerations strongly indicate that investigation of the role of repetitive sequences in nuclear structural organization in mammals is an important topic for future research. This will require a significant operational and conceptual shift. Operationally, genome-wide approaches would have to be tailored to the analysis of repetitive sequences, which represents a serious bioinformatics challenge. Conceptually, investigators should take into consideration the biological relevance of the major component of the human genome, being aware that this could potentially change the understanding of how the nuclear processes work.

ACKNOWLEDGMENTS

This work is a partial fulfillment of Valentina Casa's PhD in Molecular Medicine, Program in Neuroscience, San Raffaele University, Milano, Italy. The Gabellini laboratory is supported from the European Research Council (ERC), the Italian Epigenomics Flagship Project, the Italian Ministry of Health and the FSHD Global Research Foundation. D. Gabellini is a Dulbecco Telethon Institute Assistant Scientist.

REFERENCES

- Adamo, A., Sese, B., Boue, S., Castano, J., Paramonov, I., Barrero, M. J., et al. (2011). LSD1 regulates the balance between self-renewal and differentiation in human embryonic stem cells. *Nat. Cell Biol.* 13, 652–659.
- Allen, T. A., Von Kaenel, S., Goodrich, J. A., and Kugel, J. F. (2004). The SINE-encoded mouse B2 RNA represses mRNA transcription in response to heat shock. *Nat. Struct. Mol. Biol.* 11, 816–821.
- Ames, D., Murphy, N., Helentjaris, T., Sun, N., and Chandler, V. (2008). Comparative analyses of human single- and multilocus tandem repeats. *Genetics* 179, 1693–1704.
- Amouyal, M. (2007). Transition from DNA looping to simple binding or DNA pairing in gene regulation and replication: a matter of numbers for the cell. *Genes Genomes Genomics Glob. Sci. Books* 1, 104–112.
- Amouyal, M., Perez, N., and Rolland, S. (1998). Adjacent cooperation of proteins on DNA are not representative of long-distance interactions. *C R Acad. Sci. III* 321, 877–881.
- Bantignies, F., and Cavalli, G. (2011). Polycomb group proteins: repression in 3D. *Trends Genet.* 27, 454–464.
- Bantignies, F., Roure, V., Comet, I., Leblanc, B., Schuettengruber, B., Bonnet, J., et al. (2011). Polycomb-dependent regulatory contacts between distant Hox loci in *Drosophila*. *Cell* 144, 214–226.
- Batzner, M. A., and Deininger, P. L. (2002). Alu repeats and human genomic diversity. *Nat. Rev. Genet.* 3, 370–379.
- Bennett, E. A., Coleman, L. E., Tsui, C., Pittard, W. S., and Devine, S. E. (2004). Natural genetic variation caused by transposable elements in humans. *Genetics* 168, 933–951.
- Blackburn, E. H. (1984). The molecular structure of centromeres and telomeres. *Annu. Rev. Biochem.* 53, 163–194.
- Block, G. J., Petek, L. M., Narayanan, D., Amell, A. M., Moore, J. M., Rabaia, N. A., et al. (2012). Asymmetric bidirectional transcription from the FSHD-causing D4Z4 array modulates DUX4 production. *PLoS ONE* 7:e35532. doi: 10.1371/journal.pone.0035532
- Blower, M. D., Sullivan, B. A., and Karpen, G. H. (2002). Conserved organization of centromeric chromatin in flies and humans. *Dev. Cell* 2, 319–330.
- Bodega, B., Ramirez, G. D., Grasser, F., Cheli, S., Brunelli, S., Mora,

- M., et al. (2009). Remodeling of the chromatin structure of the facioscapulohumeral muscular dystrophy (FSHD) locus and upregulation of FSHD-related gene 1 (FRG1) expression during human myogenic differentiation. *BMC Biol.* 7, 41.
- Bollati, V., Fabris, S., Pegoraro, V., Ronchetti, D., Mosca, L., Deliliers, G. L., et al. (2009). Differential repetitive DNA methylation in multiple myeloma molecular subgroups. *Carcinogenesis* 30, 1330–1335.
- Boumil, R. M., Ogawa, Y., Sun, B. K., Huynh, K. D., and Lee, J. T. (2006). Differential methylation of Xite and CTCF sites in Tsix mirrors the pattern of X-inactivation choice in mice. *Mol. Cell. Biol.* 26, 2109–2117.
- Bourque, G., Leong, B., Vega, V. B., Chen, X., Lee, Y. L., Srinivasan, K. G., et al. (2008). Evolution of the mammalian transcription factor binding repertoire via transposable elements. *Genome Res.* 18, 1752–1762.
- Brack, C., Hiram, M., Lenhard-Schuller, R., and Tonegawa, S. (1978). A complete immunoglobulin gene is created by somatic recombination. *Cell* 15, 1–14.
- Britten, R. J. (2010). Transposable element insertions have strongly affected human evolution. *Proc. Natl. Acad. Sci. U.S.A.* 107, 19945–19948.
- Brockdorff, N., Ashworth, A., Kay, G. F., McCabe, V. M., Norris, D. P., Cooper, P. J., et al. (1992). The product of the mouse Xist gene is a 15 kb inactive X-specific transcript containing no conserved ORF and located in the nucleus. *Cell* 71, 515–526.
- Brouha, B., Schustak, J., Badge, R. M., Lutz-Prigge, S., Farley, A. H., Moran, J. V., et al. (2003). Hot L1s account for the bulk of retrotransposition in the human population. *Proc. Natl. Acad. Sci. U.S.A.* 100, 5280–5285.
- Brown, C. J., Hendrich, B. D., Rupert, J. L., Lafreniere, R. G., Xing, Y., Lawrence, J., et al. (1992). The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell* 71, 527–542.
- Buchenau, P., Hodgson, J., Strutt, H., and Arndt-Jovin, D. J. (1998). The distribution of polycomb-group proteins during cell division and development in *Drosophila* embryos: impact on models for silencing. *J. Cell Biol.* 141, 469–481.
- Buschbeck, M., Uribealago, I., Wibowo, I., Rue, P., Martin, D., Gutierrez, A., et al. (2009). The histone variant macroH2A is an epigenetic regulator of key developmental genes. *Nat. Struct. Mol. Biol.* 16, 1074–1079.
- Busturia, A., Lloyd, A., Bejarano, F., Zavortink, M., Xin, H., and Sakonju, S. (2001). The MCP silencer of the *Drosophila* Abd-B gene requires both Pleiohomeotic and GAGA factor for the maintenance of repression. *Development* 128, 2163–2173.
- Cabianca, D. S., Casa, V., Bodega, B., Xynos, A., Ginelli, E., Tanaka, Y., et al. (2012). A long ncRNA links copy number variation to a polycomb/trithorax epigenetic switch in FSHD muscular dystrophy. *Cell* 149, 819–831.
- Cabianca, D. S., and Gabellini, D. (2010). The cell biology of disease: FSHD: copy number variations on the theme of muscular dystrophy. *J. Cell Biol.* 191, 1049–1060.
- Cao, R., Wang, L., Wang, H., Xia, L., Erdjument-Bromage, H., Tempst, P., et al. (2002). Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* 298, 1039–1043.
- Chadwick, B. P. (2008). DXZ4 chromatin adopts an opposing conformation to that of the surrounding chromosome and acquires a novel inactive X-specific role involving CTCF and antisense transcripts. *Genome Res.* 18, 1259–1269.
- Chadwick, B. P. (2009). Macrosatellite epigenetics: the two faces of DXZ4 and D4Z4. *Chromosoma* 118, 675–681.
- Chadwick, B. P., and Willard, H. F. (2004). Multiple spatially distinct types of facultative heterochromatin on the human inactive X chromosome. *Proc. Natl. Acad. Sci. U.S.A.* 101, 17450–17455.
- Chao, W., Huynh, K. D., Spencer, R. J., Davidow, L. S., and Lee, J. T. (2002). CTCF, a candidate trans-acting factor for X-inactivation choice. *Science* 295, 345–347.
- Chaumeil, J., Le Baccon, P., Wutz, A., and Heard, E. (2006). A novel role for Xist RNA in the formation of a repressive nuclear compartment into which genes are recruited when silenced. *Genes Dev.* 20, 2223–2237.
- Cheutin, T., and Cavalli, G. (2012). Progressive polycomb assembly on H3K27me3 compartments generates polycomb bodies with developmentally regulated motion. *PLoS Genet.* 8:e1002465. doi: 10.1371/journal.pgen.1002465
- Choo, K. H. (1997). Centromere DNA dynamics: latent centromeres and neocentromere formation. *Am. J. Hum. Genet.* 61, 1225–1233.
- Chow, J. C., Hall, L. L., Baldry, S. E., Thorogood, N. P., Lawrence, J. B., and Brown, C. J. (2007). Inducible XIST-dependent X-chromosome inactivation in human somatic cells is reversible. *Proc. Natl. Acad. Sci. U.S.A.* 104, 10104–10109.
- Chuang, C. H., Carpenter, A. E., Fuchsova, B., Johnson, T., De Lanerolle, P., and Belmont, A. S. (2006). Long-range directional movement of an interphase chromosome site. *Curr. Biol.* 16, 825–831.
- Clapp, J., Mitchell, L. M., Bolland, D. J., Fantes, J., Corcoran, A. E., Scotting, P. J., et al. (2007). Evolutionary conservation of a coding function for D4Z4, the tandem DNA repeat mutated in facioscapulohumeral muscular dystrophy. *Am. J. Hum. Genet.* 81, 264–279.
- Cleard, F., Moshkin, Y., Karch, F., and Maeda, R. K. (2006). Probing long-distance regulatory interactions in the *Drosophila melanogaster* bithorax complex using Dam identification. *Nat. Genet.* 38, 931–935.
- Clemons, C. M., Hall, L. L., Byron, M., McNeil, J., and Lawrence, J. B. (2006). The X chromosome is organized into a gene-rich outer rim and an internal core containing silenced nongenic sequences. *Proc. Natl. Acad. Sci. U.S.A.* 103, 7688–7693.
- Comet, I., Schuettengruber, B., Sexton, T., and Cavalli, G. (2011). A chromatin insulator driving three-dimensional Polycomb response element (PRE) contacts and Polycomb association with the chromatin fiber. *Proc. Natl. Acad. Sci. U.S.A.* 108, 2294–2299.
- Conley, A. B., Miller, W. J., and Jordan, I. K. (2008). Human cis natural antisense transcripts initiated by transposable elements. *Trends Genet.* 24, 53–56.
- Coufal, N. G., Garcia-Perez, J. L., Peng, G. E., Yeo, G. W., Mu, Y., Lovci, M. T., et al. (2009). L1 retrotransposition in human neural progenitor cells. *Nature* 460, 1127–1131.
- Courtier, B., Heard, E., and Avner, P. (1995). Xce haplotypes show modified methylation in a region of the active X chromosome lying 3' to Xist. *Proc. Natl. Acad. Sci. U.S.A.* 92, 3531–3535.
- Cuddapah, S., Roh, T. Y., Cui, K., Jose, C. C., Fuller, M. T., Zhao, K., et al. (2012). A novel human polycomb binding site acts as a functional polycomb response element in *Drosophila*. *PLoS ONE* 7:e36365. doi: 10.1371/journal.pone.0036365
- Day, D. S., Luquette, L. J., Park, P. J., and Kharchenko, P. V. (2010). Estimating enrichment of repetitive elements from high-throughput sequence data. *Genome Biol.* 11, R69.
- De Koning, A. P., Gu, W., Castoe, T. A., Batzer, M. A., and Pollock, D. D. (2011). Repetitive elements may comprise over two-thirds of the human genome. *PLoS Genet.* 7:e1002384. doi: 10.1371/journal.pgen.1002384
- De Napolles, M., Mermoud, J. E., Wakao, R., Tang, Y. A., Endoh, M., Appanah, R., et al. (2004). Polycomb group proteins Ring1A/B link ubiquitylation of histone H2A to heritable gene silencing and X inactivation. *Dev. Cell* 7, 663–676.
- Deidda, G., Cacurri, S., Grisanti, P., Vigneti, E., Piazzi, N., and Felicetti, L. (1995). Physical mapping evidence for a duplicated region on chromosome 10qter showing high homology with the facioscapulohumeral muscular dystrophy locus on chromosome 4qter. *Eur. J. Hum. Genet.* 3, 155–167.
- Dejardin, J., Rappailles, A., Cuvier, O., Grimaud, C., Decoville, M., Locker, D., et al. (2005). Recruitment of *Drosophila* Polycomb group proteins to chromatin by DSP1. *Nature* 434, 533–538.
- Dixit, M., Anseau, E., Tassin, A., Winokur, S., Shi, R., Qian, H., et al. (2007). DUX4, a candidate gene of facioscapulohumeral muscular dystrophy, encodes a transcriptional activator of PITX1. *Proc. Natl. Acad. Sci. U.S.A.* 104, 18157–18162.
- Dong, K. B., Maksakova, I. A., Mohn, F., Leung, D., Appanah, R., Lee, S., et al. (2008). DNA methylation in ES cells requires the lysine methyltransferase G9a but not its catalytic activity. *EMBO J.* 27, 2691–2701.
- Donohoe, M. E., Zhang, L. F., Xu, N., Shi, Y., and Lee, J. T. (2007). Identification of a Ctcf cofactor, Yy1, for the X chromosome binary switch. *Mol. Cell* 25, 43–56.
- Duszczak, M. M., Wutz, A., Rybin, V., and Sattler, M. (2011). The Xist RNA A-repeat comprises a novel AUCG tetraloop fold and a platform for multimerization. *RNA* 17, 1973–1982.
- Eckram, M. B., Kang, K., Kim, H., and Kim, J. (2012). Retrotransposons as a major source of epigenetic variations in the mammalian genome. *Epigenetics* 7, 370–382.
- Eskeland, R., Leeb, M., Grimes, G. R., Kress, C., Boyle, S., Sproul, D., et al. (2010). Ring1B compacts

- chromatin structure and represses gene expression independent of histone ubiquitination. *Mol. Cell* 38, 452–464.
- Faulkner, G. J. (2011). Retrotransposons: mobile and mutagenic from conception to death. *FEBS Lett.* 585, 1589–1594.
- Faulkner, G. J., and Carninci, P. (2009). Altruistic functions for selfish DNA. *Cell Cycle* 8, 2895–2900.
- Faulkner, G. J., Kimura, Y., Daub, C. O., Wani, S., Plessy, C., Irvine, K. M., et al. (2009). The regulated retrotransposon transcriptome of mammalian cells. *Nat. Genet.* 41, 563–571.
- Ferraiuolo, M. A., Rousseau, M., Miyamoto, C., Shenker, S., Wang, X. Q., Nadler, M., et al. (2010). The three-dimensional architecture of Hox cluster silencing. *Nucleic Acids Res.* 38, 7472–7484.
- Feschotte, C. (2008). Transposable elements and the evolution of regulatory networks. *Nat. Rev. Genet.* 9, 397–405.
- Festenstein, R., Sharghi-Namini, S., Fox, M., Roderick, K., Tolaini, M., Norton, T., et al. (1999). Heterochromatin protein 1 modifies mammalian PEV in a dose- and chromosomal-context-dependent manner. *Nat. Genet.* 23, 457–461.
- Ficz, G., Heintzmann, R., and Arndt-Jovin, D. J. (2005). Polycomb group protein complexes exchange rapidly in living *Drosophila*. *Development* 132, 3963–3976.
- Filippova, G. N. (2008). Genetics and epigenetics of the multifunctional protein CTCF. *Curr. Top. Dev. Biol.* 80, 337–360.
- Gabellini, D., Green, M. R., and Tupler, R. (2002). Inappropriate gene activation in FSHD: a repressor complex binds a chromosomal repeat deleted in dystrophic muscle. *Cell* 110, 339–348.
- Gao, Z., Zhang, J., Bonasio, R., Strino, F., Sawai, A., Parisi, F., et al. (2012). PCGF homologs, CBX proteins, and RYBP define functionally distinct PRC1 family complexes. *Mol. Cell* 45, 344–356.
- Garrick, D., Fiering, S., Martin, D. I., and Whitelaw, E. (1998). Repeat-induced gene silencing in mammals. *Nat. Genet.* 18, 56–59.
- Gelfand, Y., Rodriguez, A., and Benson, G. (2007). TRDB – the Tandem Repeats Database. *Nucleic Acids Res.* 35, D80–D87.
- Geng, L. N., Yao, Z., Snider, L., Fong, A. P., Cech, J. N., Young, J. M., et al. (2012). DUX4 activates germline genes, retroelements, and immune mediators: implications for facioscapulohumeral dystrophy. *Dev. Cell* 22, 38–51.
- Ghildiyal, M., Seitz, H., Horwich, M. D., Li, C., Du, T., Lee, S., et al. (2008). Endogenous siRNAs derived from transposons and mRNAs in *Drosophila* somatic cells. *Science* 320, 1077–1081.
- Giacalone, J., Friedes, J., and Francke, U. (1992). A novel GC-rich human macrosatellite VNTR in Xq24 is differentially methylated on active and inactive X chromosomes. *Nat. Genet.* 1, 137–143.
- Gondo, Y., Okada, T., Matsuyama, N., Saitoh, Y., Yanagisawa, Y., and Ikeda, J. E. (1998). Human megasatellite DNA RS447, copy-number polymorphisms and interspecies conservation. *Genomics* 54, 39–49.
- Grimaud, C., Bantignies, F., Pal-Bhadra, M., Ghana, P., Bhadra, U., and Cavalli, G. (2006). RNAi components are required for nuclear clustering of Polycomb group response elements. *Cell* 124, 957–971.
- Guenatri, M., Bailly, D., Maison, C., and Almouzni, G. (2004). Mouse centric and pericentric satellite repeats form distinct functional heterochromatin. *J. Cell Biol.* 166, 493–505.
- Guil, S., Soler, M., Portela, A., Carrere, J., Fonalleras, E., Gomez, A., et al. (2012). Intronic RNAs mediate EZH2 regulation of epigenetic targets. *Nat. Struct. Mol. Biol.* 19, 664–670.
- Gupta, R. A., Shah, N., Wang, K. C., Kim, J., Horlings, H. M., Wong, D. J., et al. (2010). Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* 464, 1071–1076.
- Hall, L. L., and Lawrence, J. B. (2003). The cell biology of a novel chromosomal RNA: chromosome painting by XIST/Xist RNA initiates a remodeling cascade. *Semin. Cell Dev. Biol.* 14, 369–378.
- Hall, L. L., and Lawrence, J. B. (2010). XIST RNA and architecture of the inactive X chromosome: implications for the repeat genome. *Cold Spring Harb. Symp. Quant. Biol.* 75, 345–356.
- Hall, L. L., Smith, K. P., Byron, M., and Lawrence, J. B. (2006). Molecular anatomy of a speckle. *Anat. Rec. A Discov. Mol. Cell. Evol. Biol.* 288, 664–675.
- Heard, E., and Distech, C. M. (2006). Dosage compensation in mammals: fine-tuning the expression of the X chromosome. *Genes Dev.* 20, 1848–1867.
- Hendrich, B. D., Plenge, R. M., and Willard, H. F. (1997). Identification and characterization of the human XIST gene promoter: implications for models of X chromosome inactivation. *Nucleic Acids Res.* 25, 2661–2671.
- Hiragami, K., and Festenstein, R. (2005). Heterochromatin protein 1, a pervasive controlling influence. *Cell. Mol. Life Sci.* 62, 2711–2726.
- Horvath, J. E., Sheedy, C. B., Merrett, S. L., Diallo, A. B., Swofford, D. L., Program, N. C. S., et al. (2011). Comparative analysis of the primate X-inactivation center region and reconstruction of the ancestral primate XIST locus. *Genome Res.* 21, 850–862.
- Hua-Van, A., Le Rouzic, A., Boutin, T. S., Filee, J., and Capi, P. (2011). The struggle for life of the genome's selfish architects. *Biol. Direct* 6, 19.
- Igarashi, S., Suzuki, H., Niinuma, T., Shimizu, H., Nojima, M., Iwaki, H., et al. (2010). A novel correlation between LINE-1 hypomethylation and the malignancy of gastrointestinal stromal tumors. *Clin. Cancer Res.* 16, 5114–5123.
- Iskow, R. C., McCabe, M. T., Mills, R. E., Torene, S., Pittard, W. S., Neuwald, A. F., et al. (2010). Natural mutagenesis of human genomes by endogenous retrotransposons. *Cell* 141, 1253–1261.
- Jeon, Y., and Lee, J. T. (2011). YY1 tethers Xist RNA to the inactive X nucleation center. *Cell* 146, 119–133.
- Jiang, G., Yang, F., Van Overveld, P. G., Vandarayanan, V., Van Der Maarel, S., and Ehrlich, M. (2003). Testing the position-effect variation hypothesis for facioscapulohumeral muscular dystrophy by analysis of histone modification and gene expression in subtelomeric 4q. *Hum. Mol. Genet.* 12, 2909–2921.
- Jurka, J., Kapitonov, V. V., Kohany, O., and Jurka, M. V. (2007). Repetitive sequences in complex genomes: structure and evolution. *Annu. Rev. Genomics Hum. Genet.* 8, 241–259.
- Kaneko, H., Dridi, S., Tarallo, V., Gelfand, B. D., Fowler, B. J., Cho, W. G., et al. (2011). DICER1 deficit induces Alu RNA toxicity in age-related macular degeneration. *Nature* 471, 325–330.
- Kanhere, A., Viiri, K., Araujo, C. C., Rasaiyaah, J., Bouwman, R. D., Whyte, W. A., et al. (2010). Short RNAs are transcribed from repressed polycomb target genes and interact with polycomb repressive complex-2. *Mol. Cell* 38, 675–688.
- Kano, H., Godoy, I., Courtney, C., Vetter, M. R., Gerton, G. L., Ostertag, E. M., et al. (2009). L1 retrotransposition occurs mainly in embryogenesis and creates somatic mosaicism. *Genes Dev.* 23, 1303–1312.
- Kato, Y., Kaneda, M., Hata, K., Kumaki, K., Hisano, M., Kohara, Y., et al. (2007). Role of the Dnmt3 family in de novo methylation of imprinted and repetitive sequences during male germ cell development in the mouse. *Hum. Mol. Genet.* 16, 2272–2280.
- Kazazian, H. H. Jr., Wong, C., Youssoufian, H., Scott, A. F., Phillips, D. G., and Antonarakis, S. E. (1988). Haemophilia A resulting from de novo insertion of L1 sequences represents a novel mechanism for mutation in man. *Nature* 332, 164–166.
- Khalil, A. M., Guttman, M., Huarde, M., Garber, M., Raj, A., Rivea Morales, D., et al. (2009). Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc. Natl. Acad. Sci. U.S.A.* 106, 11667–11672.
- Kidwell, M. G., and Lisch, D. R. (2000). Transposable elements and host genome evolution. *Trends Ecol. Evol.* 15, 95–99.
- Kigami, D., Minami, N., Takayama, H., and Imai, H. (2003). MuERV-L is one of the earliest transcribed genes in mouse one-cell embryos. *Biol. Reprod.* 68, 651–654.
- Kim, P. M., Lam, H. Y., Urban, A. E., Korbel, J. O., Affourtit, J., Grubert, F., et al. (2008). Analysis of copy number variants and segmental duplications in the human genome: Evidence for a change in the process of formation in recent evolutionary history. *Genome Res.* 18, 1865–1874.
- Kogi, M., Fukushige, S., Lefevre, C., Hadano, S., and Ikeda, J. E. (1997). A novel tandem repeat sequence located on human chromosome 4p: isolation and characterization. *Genomics* 42, 278–283.
- Kohlmaier, A., Savarese, F., Lachner, M., Martens, J., Jenuwein, T., and Wutz, A. (2004). A chromosomal memory triggered by Xist regulates histone methylation in X inactivation. *PLoS Biol.* 2:e171. doi: 10.1371/journal.pbio.0020171
- Kosak, S. T., Skok, J. A., Medina, K. L., Riblet, R., Le Beau, M. M., Fisher, A. G., et al. (2002). Subnuclear compartmentalization of immunoglobulin loci during lymphocyte development. *Science* 296, 158–162.
- Ku, M., Koche, R. P., Rheinbay, E., Mendenhall, E. M., Endoh, M., Mikkelsen, T. S., et al. (2008). Genomewide analysis of PRC1 and PRC2 occupancy identifies two classes of bivalent domains.

- PLoS Genet.* 4:e1000242. doi: 10.1371/journal.pgen.1000242
- Kuzmichev, A., Jenuwein, T., Tempst, P., and Reinberg, D. (2004). Different EZH2-containing complexes target methylation of histone H1 or nucleosomal histone H3. *Mol. Cell* 14, 183–193.
- Kuzmichev, A., Margueron, R., Vaquero, A., Preissner, T. S., Scher, M., Kirmizis, A., et al. (2005). Composition and histone substrates of polycomb repressive group complexes change during cellular differentiation. *Proc. Natl. Acad. Sci. U.S.A.* 102, 1859–1864.
- Kuzmichev, A., Nishioka, K., Erdjument-Bromage, H., Tempst, P., and Reinberg, D. (2002). Histone methyltransferase activity associated with a human multiprotein complex containing the Enhancer of Zeste protein. *Genes Dev.* 16, 2893–2905.
- Lamond, A. I., and Spector, D. L. (2003). Nuclear speckles: a model for nuclear organelles. *Nat. Rev. Mol. Cell Biol.* 4, 605–612.
- Lancot, C., Cheutin, T., Cremer, M., Cavalli, G., and Cremer, T. (2007). Dynamic genome architecture in the nuclear space: regulation of gene expression in three dimensions. *Nat. Rev. Genet.* 8, 104–115.
- Landeira, D., Sauer, S., Poot, R., Dvorkina, M., Mazzarella, L., Jorgensen, H. F., et al. (2010). Jarid2 is a PRC2 component in embryonic stem cells required for multi-lineage differentiation and recruitment of PRC1 and RNA Polymerase II to developmental regulators. *Nat. Cell Biol.* 12, 618–624.
- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., et al. (2001). Initial sequencing and analysis of the human genome. *Nature* 409, 860–921.
- Lanzuolo, C., Roure, V., Dekker, J., Bantignies, F., and Orlando, V. (2007). Polycomb response elements mediate the formation of chromosome higher-order structures in the bithorax complex. *Nat. Cell Biol.* 9, 1167–1174.
- Lee, E., Iskow, R., Yang, L., Gokcumen, O., Haseley, P., Luquette, L. J., et al. (2012). Landscape of somatic retrotransposition in human cancers. *Science* 337, 967–971.
- Leeb, M., Pasini, D., Novatchkova, M., Jaritz, M., Helin, K., and Wutz, A. (2010). Polycomb complexes act redundantly to repress genomic repeats and genes. *Genes Dev.* 24, 265–276.
- Leeb, M., and Wutz, A. (2007). Ring1B is crucial for the regulation of developmental control genes and PRC1 proteins but not X inactivation in embryonic cells. *J. Cell Biol.* 178, 219–229.
- Leidenroth, A., and Hewitt, J. E. (2010). A family history of DUX4, phylogenetic analysis of DUXA, B, C and Duxbl reveals the ancestral DUX gene. *BMC Evol. Biol.* 10, 364.
- Lemmers, R. J., Van Der Vliet, P. J., Klooster, R., Sacconi, S., Camano, P., Dauwerse, J. G., et al. (2010). A unifying genetic model for facioscapulohumeral muscular dystrophy. *Science* 329, 1650–1653.
- Li, G., Margueron, R., Ku, M., Chambon, P., Bernstein, B. E., and Reinberg, D. (2010). Jarid2 and PRC2, partners in regulating gene expression. *Genes Dev.* 24, 368–380.
- Li, H. B., Muller, M., Bahechar, I. A., Kyrchanova, O., Ohno, K., Georgiev, P., et al. (2011). Insulators, not Polycomb response elements, are required for long-range interactions between Polycomb targets in *Drosophila melanogaster*. *Mol. Cell Biol.* 31, 616–625.
- Liang, G., Chan, M. F., Tomigahara, Y., Tsai, Y. C., Gonzales, F. A., Li, E., et al. (2002). Cooperativity between DNA methyltransferases in the maintenance methylation of repetitive elements. *Mol. Cell Biol.* 22, 480–491.
- Ling, J. Q., Li, T., Hu, J. F., Vu, T. H., Chen, H. L., Qiu, X. W., et al. (2006). CTCF mediates interchromosomal colocalization between Igf2/H19 and Wsb1/Nf1. *Science* 312, 269–272.
- Lo, A. W., Craig, J. M., Saffery, R., Kalitsis, P., Irvine, D. V., Earle, E., et al. (2001). A 330 kb CENP-A binding domain and altered replication timing at a human neocentromere. *EMBO J.* 20, 2087–2096.
- Lopez Castel, A., Cleary, J. D., and Pearson, C. E. (2010). Repeat instability as the basis for human diseases and as a potential target for therapy. *Nat. Rev. Mol. Cell Biol.* 11, 165–170.
- Lunyak, V. V., Prefontaine, G. G., Nunez, E., Cramer, T., Ju, B. G., Ohgi, K. A., et al. (2007). Developmentally regulated activation of a SINE B2 repeat as a domain boundary in organogenesis. *Science* 317, 248–251.
- Lyle, R., Wright, T. J., Clark, L. N., and Hewitt, J. E. (1995). The FSHD-associated repeat, D4Z4, is a member of a dispersed family of homeobox-containing repeats, subsets of which are clustered on the short arms of the acrocentric chromosomes. *Genomics* 28, 389–397.
- Macfarlan, T. S., Gifford, W. D., Agarwal, S., Driscoll, S., Lettieri, K., Wang, J., et al. (2011). Endogenous retroviruses and neighboring genes are coordinately repressed by LSD1/KDM1A. *Genes Dev.* 25, 594–607.
- Macfarlan, T. S., Gifford, W. D., Driscoll, S., Lettieri, K., Rowe, H. M., Bonanomi, D., et al. (2012). Embryonic stem cell potency fluctuates with endogenous retrovirus activity. *Nature* 487, 57–63.
- Maksakova, I. A., Mager, D. L., and Reiss, D. (2008). Keeping active endogenous retroviral-like elements in check: the epigenetic perspective. *Cell. Mol. Life Sci.* 65, 3329–3347.
- Martens, J. H., O'Sullivan, R. J., Braunschweig, U., Opravil, S., Radolf, M., Steinlein, P., et al. (2005). The profile of repeat-associated histone lysine methylation states in the mouse epigenome. *EMBO J.* 24, 800–812.
- Masny, P. S., Bengtsson, U., Chung, S. A., Martin, J. H., Van Engelen, B., Van Der Maarel, S. M., et al. (2004). Localization of 4q35.2 to the nuclear periphery: is FSHD a nuclear envelope disease? *Hum. Mol. Genet.* 13, 1857–1871.
- Matharu, N. K., Hussain, T., Sankaranarayanan, R., and Mishra, R. K. (2010). Vertebrate homologue of *Drosophila* GAGA factor. *J. Mol. Biol.* 400, 434–447.
- McLaughlin, C. R., and Chadwick, B. P. (2011). Characterization of DXZ4 conservation in primates implies important functional roles for CTCF binding, array expression and tandem repeat organization on the X chromosome. *Genome Biol.* 12, R37.
- Meaburn, K. J., and Misteli, T. (2007). Cell biology: chromosome territories. *Nature* 445, 379–781.
- Memili, E., Hong, Y. K., Kim, D. H., Ontiveros, S. D., and Strauss, W. M. (2001). Murine Xist RNA isoforms are different at their 3' ends: a role for differential polyadenylation. *Gene* 266, 131–137.
- Mendenhall, E. M., Koche, R. P., Truong, T., Zhou, V. W., Issac, B., Chi, A. S., et al. (2010). GC-rich sequence elements recruit PRC2 in mammalian ES cells. *PLoS Genet.* 6:e1001244. doi: 10.1371/journal.pgen.1001244
- Messmer, S., Franke, A., and Paro, R. (1992). Analysis of the functional role of the Polycomb chromo domain in *Drosophila melanogaster*. *Genes Dev.* 6, 1241–1254.
- Mihaly, J., Mishra, R. K., and Karch, F. (1998). A conserved sequence motif in Polycomb-response elements. *Mol. Cell* 1, 1065–1066.
- Mikkelsen, T. S., Ku, M., Jaffe, D. B., Issac, B., Lieberman, E., Giannoukos, G., et al. (2007). Genome-wide maps of chromatin state in pluripotent and lineage-committed cells. *Nature* 448, 553–560.
- Mills, R. E., Bennett, E. A., Iskow, R. C., and Devine, S. E. (2007). Which transposable elements are active in the human genome? *Trends Genet.* 23, 183–191.
- Mishra, R. K., Mihaly, J., Barges, S., Spierer, A., Karch, F., Hagstrom, K., et al. (2001). The iab-7 polycomb response element maps to a nucleosome-free region of chromatin and requires both GAGA and pleiohomeotic for silencing activity. *Mol. Cell Biol.* 21, 1311–1318.
- Morey, L., and Helin, K. (2010). Polycomb group protein-mediated repression of transcription. *Trends Biochem. Sci.* 35, 323–332.
- Morris, C. A., and Moazed, D. (2007). Centromere assembly and propagation. *Cell* 128, 647–650.
- Moseley, S. C., Rizkallah, R., Tremblay, D. C., Anderson, B. R., Hurt, M. M., and Chadwick, B. P. (2012). YY1 associates with the macrosatellite DXZ4 on the inactive X chromosome and binds with CTCF to a hypomethylated form in some male carcinomas. *Nucleic Acids Res.* 40, 1596–1608.
- Muller, J., and Kassiss, J. A. (2006). Polycomb response elements and targeting of Polycomb group proteins in *Drosophila*. *Curr. Opin. Genet. Dev.* 16, 476–484.
- Muotri, A. R., Chu, V. T., Marchetto, M. C. N., Deng, W., Moran, J. V., and Gage, F. H. (2005). Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. *Nature* 435, 903–910.
- Muotri, A. R., Marchetto, M. C. N., Coufal, N. G., Oefner, R., Yeo, G., Nakashima, K., et al. (2010). L1 retrotransposition in neurons is modulated by MeCP2. *Nature* 468, 443–446.
- Myers, R. M., Stamatoyannopoulos, J. S., Dunham, I., Hardison, R. C., Bernstein, B. E., Gingeras, T. R., et al. (2011). A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol.* 9:e1001046. doi: 10.1371/journal.pbio.1001046
- Nguembor, M. V., and Gabellini, D. (2010). In junk we trust: repetitive DNA, epigenetics and facioscapulohumeral muscular dystrophy. *Epigenomics* 2, 271–287.
- Nesterova, T. B., Slobodyanyuk, S. Y., Elisaphenko, E. A., Shevchenko, A. I., Johnston, C., Pavlova, M.

- E., et al. (2001). Characterization of the genomic Xist locus in rodents reveals conservation of overall gene structure and tandem repeats but rapid evolution of unique sequence. *Genome Res.* 11, 833–849.
- Norris, J., Fan, D., Aleman, C., Marks, J. R., Futreal, P. A., Wiseman, R. W., et al. (1995). Identification of a new subclass of Alu DNA repeats which can function as estrogen receptor-dependent transcriptional enhancers. *J. Biol. Chem.* 270, 22777–22782.
- Osborne, C. S., Chakalova, L., Brown, K. E., Carter, D., Horton, A., Debrand, E., et al. (2004). Active genes dynamically colocalize to shared sites of ongoing transcription. *Nat. Genet.* 36, 1065–1071.
- Ottaviani, A., Rival-Gervier, S., Boussouar, A., Foerster, A. M., Rondier, D., Sacconi, S., et al. (2009). The D4Z4 macrosatellite repeat acts as a CTCF and A-type lamins-dependent insulator in facio-scapulo-humeral dystrophy. *PLoS Genet.* 5:e1000394. doi: 10.1371/journal.pgen.1000394
- Pasini, D., Cloos, P. A., Walfridsson, J., Olsson, L., Bukowski, J. P., Johansen, J. V., et al. (2010). JARID2 regulates binding of the Polycomb repressive complex 2 to target genes in ES cells. *Nature* 464, 306–310.
- Peaston, A. E., Evsikov, A. V., Graber, J. H., De Vries, W. N., Holbrook, A. E., Solter, D., et al. (2004). Retrotransposons regulate host genes in mouse oocytes and preimplantation embryos. *Dev. Cell* 7, 597–606.
- Peng, J. C., Valouev, A., Swigut, T., Zhang, J., Zhao, Y., Sidow, A., et al. (2009). Jarid2/Jumonji coordinates control of PRC2 enzymatic activity and target gene occupancy in pluripotent cells. *Cell* 139, 1290–1302.
- Peters, A. H., Kubicek, S., Mechtler, K., O'Sullivan, R. J., Derijck, A. A., Perez-Burgos, L., et al. (2003). Partitioning and plasticity of repressive histone methylation states in mammalian chromatin. *Mol. Cell* 12, 1577–1589.
- Peters, A. H., O'Carroll, D., Scherthan, H., Mechtler, K., Sauer, S., Schofer, C., et al. (2001). Loss of the Suv39h histone methyltransferase impairs mammalian heterochromatin and genome stability. *Cell* 107, 323–337.
- Pheasant, M., and Mattick, J. S. (2007). Raising the estimate of functional human sequences. *Genome Res.* 17, 1245–1253.
- Rauch, T. A., Zhong, X., Wu, X., Wang, M., Kernstine, K. H., Wang, Z., et al. (2008). High-resolution mapping of DNA hypermethylation and hypomethylation in lung cancer. *Proc. Natl. Acad. Sci. U.S.A.* 105, 252–257.
- Ren, X., Vincenz, C., and Kerppola, T. K. (2008). Changes in the distributions and dynamics of polycomb repressive complexes during embryonic stem cell differentiation. *Mol. Cell. Biol.* 28, 2884–2895.
- Rinn, J. L., Kertesz, M., Wang, J. K., Squazzo, S. L., Xu, X., Bruggmann, S. A., et al. (2007). Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* 129, 1311–1323.
- Rudd, M. K., Wray, G. A., and Willard, H. F. (2006). The evolutionary dynamics of alpha-satellite. *Genome Res.* 16, 88–96.
- Sarma, K., Levasseur, P., Aristarkhov, A., and Lee, J. T. (2010). Locked nucleic acids (LNAs) reveal sequence requirements and kinetics of Xist RNA localization to the X chromosome. *Proc. Natl. Acad. Sci. U.S.A.* 107, 22196–22201.
- Schoeffner, S., Sengupta, A. K., Kubicek, S., Mechtler, K., Spahn, L., Koseki, H., et al. (2006). Recruitment of PRC1 function at the initiation of X inactivation independent of PRC2 and silencing. *EMBO J.* 25, 3110–3122.
- Schueler, M. G., Higgins, A. W., Rudd, M. K., Gustashaw, K., and Willard, H. F. (2001). Genomic and genetic definition of a functional human centromere. *Science* 294, 109–115.
- Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B., and Cavalli, G. (2007). Genome regulation by polycomb and trithorax proteins. *Cell* 128, 735–745.
- Schuettengruber, B., Martinez, A. M., Iovino, N., and Cavalli, G. (2011). Trithorax group proteins: switching genes on and keeping them active. *Nat. Rev. Mol. Cell Biol.* 12, 799–814.
- Shen, S., Lin, L., Cai, J. J., Jiang, P., Kenkel, E. J., Stroik, M. R., et al. (2011). Widespread establishment and regulatory impact of Alu exons in human genes. *Proc. Natl. Acad. Sci. U.S.A.* 108, 2837–2842.
- Shen, X., Kim, W., Fujiwara, Y., Simon, M. D., Liu, Y., Mysliwiec, M. R., et al. (2009). Jumonji modulates polycomb activity and self-renewal versus differentiation of stem cells. *Cell* 139, 1303–1314.
- Shopland, L. S., Johnson, C. V., Byron, M., McNeil, J., and Lawrence, J. B. (2003). Clustering of multiple specific genes and gene-rich R-bands around SC-35 domains: evidence for local euchromatic neighborhoods. *J. Cell Biol.* 162, 981–990.
- Simeonova, I., Lejour, V., Bardot, B., Bouarich-Bourimi, R., Morin, A., Fang, M., et al. (2012). Fuzzy tandem repeats containing p53 response elements may define species-specific p53 target genes. *PLoS Genet.* 8:e1002731. doi: 10.1371/journal.pgen.1002731
- Simons, C., Pheasant, M., Makunin, I. V., and Mattick, J. S. (2006). Transposon-free regions in mammalian genomes. *Genome Res.* 16, 164–172.
- Sing, A., Pannell, D., Karaïskakis, A., Sturgeon, K., Djabali, M., Ellis, J., et al. (2009). A vertebrate Polycomb response element governs segmentation of the posterior hindbrain. *Cell* 138, 885–897.
- Slotkin, R. K., and Martienssen, R. (2007). Transposable elements and the epigenetic regulation of the genome. *Nat. Rev. Genet.* 8, 272–285.
- Smit, A. F. A., Hubley, R., and Green, P. (1996–2004). *RepeatMasker Open-3.0*. Available online at: <http://www.repeatmasker.org>. (Accessed date: September 14, 2012).
- Snider, L., Asawachaicharn, A., Tyler, A. E., Geng, L. N., Petek, L. M., Maves, L., et al. (2009). RNA transcripts, miRNA-sized fragments and proteins produced from D4Z4 units: new candidates for the pathophysiology of facioscapulohumeral dystrophy. *Hum. Mol. Genet.* 18, 2414–2430.
- Snider, L., Geng, L. N., Lemmers, R. J., Kyba, M., Ware, C. B., Nelson, A. M., et al. (2010). Facioscapulohumeral dystrophy: incomplete suppression of a retrotransposed gene. *PLoS Genet.* 6:e1001181. doi: 10.1371/journal.pgen.1001181
- Solyom, S., and Kazazian, H. H. Jr. (2012). Mobile elements in the human genome: implications for disease. *Genome Med.* 4, 12.
- Soudais, C., Bielinska, M., Heikineimo, M., Macarthur, C. A., Narita, N., Saffitz, J. E., et al. (1995). Targeted mutagenesis of the transcription factor GATA-4 gene in mouse embryonic stem cells disrupts visceral endoderm differentiation *in vitro*. *Development* 121, 3877–3888.
- Speck, M. (2001). Antisense promoter of human L1 retrotransposon drives transcription of adjacent cellular genes. *Mol. Cell. Biol.* 21, 1973–1985.
- Spilianakis, C. G., Lalioti, M. D., Town, T., Lee, G. R., and Flavell, R. A. (2005). Interchromosomal associations between alternatively expressed loci. *Nature* 435, 637–645.
- Su, J., Shao, X., Liu, H., Liu, S., Wu, Q., and Zhang, Y. (2012). Genome-wide dynamic changes of DNA methylation of repetitive elements in human embryonic stem cells and fetal fibroblasts. *Genomics* 99, 10–17.
- Svoboda, P., Stein, P., Anger, M., Bernstein, E., Hannon, G. J., and Schultz, R. M. (2004). RNAi and expression of retrotransposons MuERV-L and IAP in preimplantation mouse embryos. *Dev. Biol.* 269, 276–285.
- Taft, R. J., Pheasant, M., and Mattick, J. S. (2007). The relationship between non-protein-coding DNA and eukaryotic complexity. *Bioessays* 29, 288–299.
- Tam, R., Smith, K. P., and Lawrence, J. B. (2004). The 4q subtelomere harboring the FSHD locus is specifically anchored with peripheral heterochromatin unlike most human telomeres. *J. Cell Biol.* 167, 269–279.
- Tanay, A., O'Donnell, A. H., Damelin, M., and Bestor, T. H. (2007). Hyperconserved CpG domains underlie Polycomb-binding sites. *Proc. Natl. Acad. Sci. U.S.A.* 104, 5521–5526.
- Tavares, L., Dimitrova, E., Oxley, D., Webster, J., Poot, R., Demmers, J., et al. (2012). RYBP-PRC1 complexes mediate H2A ubiquitylation at polycomb target sites independently of PRC2 and H3K27me3. *Cell* 148, 664–678.
- Terranova, R., Yokobayashi, S., Stadler, M. B., Otte, A. P., Van Lohuizen, M., Orkin, S. H., et al. (2008). Polycomb group proteins Ezh2 and Rnf2 direct genomic contraction and imprinted repression in early mouse embryos. *Dev. Cell* 15, 668–679.
- Tian, D., Sun, S., and Lee, J. T. (2010). The long noncoding RNA, Jpx, is a molecular switch for X chromosome inactivation. *Cell* 143, 390–403.
- Ting, D. T., Lipson, D., Paul, S., Brannigan, B. W., Akhavanfard, S., Coffman, E. J., et al. (2011). Aberrant overexpression of satellite repeats in pancreatic and other epithelial cancers. *Science* 331, 593–596.
- Tiwari, V. K., McGarvey, K. M., Licchesi, J. D., Ohm, J. E., Herman, J. G., Schubeler, D., et al. (2008). PcG proteins, DNA methylation, and gene repression by chromatin looping. *PLoS Biol.* 6:e306. doi: 10.1371/journal.pbio.0060306

- Tolhuis, B., Blom, M., Kerkhoven, R. M., Pagie, L., Teunissen, H., Nieuwland, M., et al. (2011). Interactions among Polycomb domains are guided by chromosome architecture. *PLoS Genet.* 7:e1001343. doi: 10.1371/journal.pgen.1001343
- Tremblay, D. C., Alexander, G. Jr., Moseley, S., and Chadwick, B. P. (2010). Expression, tandem repeat copy number variation and stability of four macrosatellite arrays in the human genome. *BMC Genomics* 11, 632.
- Tremblay, D. C., Moseley, S., and Chadwick, B. P. (2011). Variation in array size, monomer composition and expression of the macrosatellite DXZ4. *PLoS ONE* 6: e18969. doi: 10.1371/journal.pone.0018969
- Tyekucheva, S., Yolken, R. H., McCombie, W. R., Parla, J., Kramer, M., Wheelan, S. J., et al. (2011). Establishing the baseline level of repetitive element expression in the human cortex. *BMC Genomics* 12, 495.
- Van Deutekom, J. C., Wijmenga, C., Van Tienhoven, E. A., Gruter, A. M., Hewitt, J. E., Padberg, G. W., et al. (1993). FSHD associated DNA rearrangements are due to deletions of integral copies of a 3.2 kb tandemly repeated unit. *Hum. Mol. Genet.* 2, 2037–2042.
- Van Overveld, P. G., Lemmers, R. J., Sandkuijl, L. A., Enthoven, L., Winokur, S. T., Bakels, F., et al. (2003). Hypomethylation of D4Z4 in 4q-linked and non-4q-linked facioscapulohumeral muscular dystrophy. *Nat. Genet.* 35, 315–317.
- Walsh, C. P., Chaillet, J. R., and Bestor, T. H. (1998). Transcription of IAP endogenous retroviruses is constrained by cytosine methylation. *Nat. Genet.* 20, 116–117.
- Wang, H., Wang, L., Erdjument-Bromage, H., Vidal, M., Tempst, P., Jones, R. S., et al. (2004). Role of histone H2A ubiquitination in Polycomb silencing. *Nature* 431, 873–878.
- Wang, T., Zeng, J., Lowe, C. B., Sellers, R. G., Salama, S. R., Yang, M., et al. (2007). Species-specific endogenous retroviruses shape the transcriptional network of the human tumor suppressor protein p53. *Proc. Natl. Acad. Sci. U.S.A.* 104, 18613–18618.
- Warburton, P. E., Haaf, T., Gosden, J., Lawson, D., and Willard, H. F. (1996). Characterization of a chromosome-specific chimpanzee alpha satellite subset: evolutionary relationship to subsets on human chromosomes. *Genomics* 33, 220–228.
- Warburton, P. E., Hasson, D., Guillem, F., Lescale, C., Jin, X., and Abrusan, G. (2008). Analysis of the largest tandemly repeated DNA families in the human genome. *BMC Genomics* 9, 533.
- Watanabe, T., Totoki, Y., Toyoda, A., Kaneda, M., Kuramochi-Miyagawa, S., Obata, Y., et al. (2008). Endogenous siRNAs from naturally formed dsRNAs regulate transcripts in mouse oocytes. *Nature* 453, 539–543.
- Waterston, R. H., Lindblad-Toh, K., Birney, E., Rogers, J., Abril, J. F., Agarwal, P., et al. (2002). Initial sequencing and comparative analysis of the mouse genome. *Nature* 420, 520–562.
- Wijmenga, C., Frants, R. R., Brouwer, O. F., Moerer, P., Weber, J. L., and Padberg, G. W. (1990). Location of facioscapulohumeral muscular dystrophy gene on chromosome 4. *Lancet* 336, 651–653.
- Wijmenga, C., Hewitt, J. E., Sandkuijl, L. A., Clark, L. N., Wright, T. J., Dauwerse, H. G., et al. (1992). Chromosome 4q DNA rearrangements associated with facioscapulohumeral muscular dystrophy. *Nat. Genet.* 2, 26–30.
- Wijmenga, C., Padberg, G. W., Moerer, P., Wiegant, J., Liem, L., Brouwer, O. F., et al. (1991). Mapping of facioscapulohumeral muscular dystrophy gene to chromosome 4q35-qter by multipoint linkage analysis and in situ hybridization. *Genomics* 9, 570–575.
- Winokur, S. T., Bengtsson, U., Vargas, J. C., Wasmuth, J. J., Altherr, M. R., Weiffenbach, B., et al. (1996). The evolutionary distribution and structural organization of the homeobox-containing repeat D4Z4 indicates a functional role for the ancestral copy in the FSHD region. *Hum. Mol. Genet.* 5, 1567–1575.
- Woo, C. J., Kharchenko, P. V., Daheron, L., Park, P. J., and Kingston, R. E. (2010). A region of the human HOXD cluster that confers polycomb-group responsiveness. *Cell* 140, 99–110.
- Wutz, A., Rasmussen, T. P., and Jaenisch, R. (2002). Chromosomal silencing and localization are mediated by different domains of Xist, R. *Nat. Genet.* 30, 167–174.
- Yen, Z. C., Meyer, I. M., Karalic, S., and Brown, C. J. (2007). A cross-species comparison of X-chromosome inactivation in Eutheria. *Genomics* 90, 453–463.
- Yoda, K., Ando, S., Morishita, S., Houmura, K., Hashimoto, K., Takeyasu, K., et al. (2000). Human centromere protein A (CENP-A) can replace histone H3 in nucleosome reconstitution *in vitro*. *Proc. Natl. Acad. Sci. U.S.A.* 97, 7266–7271.
- Zeng, W., De Greef, J. C., Chen, Y. Y., Chien, R., Kong, X., Gregson, H. C., et al. (2009). Specific loss of histone H3 lysine 9 trimethylation and HP1gamma/cohesin binding at D4Z4 repeats is associated with facioscapulohumeral dystrophy (FSHD). *PLoS Genet.* 5:e1000559. doi: 10.1371/journal.pgen.1000559
- Zhang, X. H., and Chasin, L. A. (2006). Comparison of multiple vertebrate genomes reveals the birth and evolution of human exons. *Proc. Natl. Acad. Sci. U.S.A.* 103, 13427–13432.
- Zhao, J., Sun, B. K., Erwin, J. A., Song, J. J., and Lee, J. T. (2008). Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* 322, 750–756.
- Zhu, Q., Pao, G. M., Huynh, A. M., Suh, H., Tonnu, N., Nederlof, P. M., et al. (2011). BRCA1 tumour suppression occurs via heterochromatin-mediated silencing. *Nature* 477, 179–184.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received: 06 July 2012; accepted: 17 September 2012; published online: 08 October 2012.

Citation: Casa V and Gabellini D (2012) A repetitive elements perspective in Polycomb epigenetics. *Front. Gene.* 3:199. doi: 10.3389/fgene.2012.00199

This article was submitted to *Frontiers in Epigenomics*, a specialty of *Frontiers in Genetics*.

Copyright © 2012 Casa and Gabellini. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in other forums, provided the original authors and source are credited and subject to any copyright notices concerning any third-party graphics etc.